INSTITUT NATIONAL DE RECHERCHE SUR LES TRANSPORTS ET LEUR SÉCURITÉ

Mehdi DANECH-PAJOUH Maurice ARON

ATHENA

PRÉVISION A COURT TERME DU TRAFIC SUR UNE SECTION DE ROUTE

Rapport INRETS N° 177 Avril 1994

Mehdi DANECH-PAJOUH, chargé de recherche au Département d'Analyse et Régulation du Trafic
Maurice ARON chargé de recherche au Département Mathématiques Appliquées et Intelligence Artificielle

INSTITUT NATIONAL DE RECHERCHE SUR LES TRANSPORTS ET LEUR SÉCURITÉ 2, avenue du Général Malleret-Joinville, F-94114 ARCUEIL CEDEX
Téléphone : # (33) (1) 47 40 70 00-Télécopie : # (33) (1) 45 47 56 06
ISBN 2-85782-398-3

Rapport INRTS n° 177

FICHE RESULTAT

ATHENA
PREVISION A COURT TERME DU TRAFIC SUR UNE SECTION DE ROUTE

M. DANECH-PAJOUH M. ARON

OBJECTIF :

L'objectif principal d' ATHENA est de prédire, pour un horizon court (une demi-heure à deux heures), le trafic d'une section de route. Ce type de prédiction, qui est plutôt destiné aux gestionnaires du trafic qu'aux usagers de la route, exige, pour être utilisable, une précision dans la prédiction, dépassant celle des gestionnaires.

CONCEPTION DU MODELE

Toute modélisation mathématique impose un certain nombre d'hypothèses. La première hypothèse d'ATHENA est que le trafic de l'avenir proche (une demi-heure à deux heures) dépend du trafic du moment et aussi du passé proche, la seconde est que cette dépendance peut se définir par un modèle linéaire constitué à partir des données historiques. Cette hypothèse de linéarité est atténuée par la constitution d'une série de modèles, chaque modèle traduisant un certain type de trafic. Ainsi l'hypothèse de linéarité de base se réalise dans un processus non linéaire.

La méthode comprend trois étapes principales :

- l'analyse typologique des données historiques,
- la modélisation par type du trafic faisant intervenir, à la fois les considérations temporelles et spatiales,
- la prédiction cherche d'abord, pour une situation donnée, à identifier le type du trafic et ensuite applique le modèle correspondant.

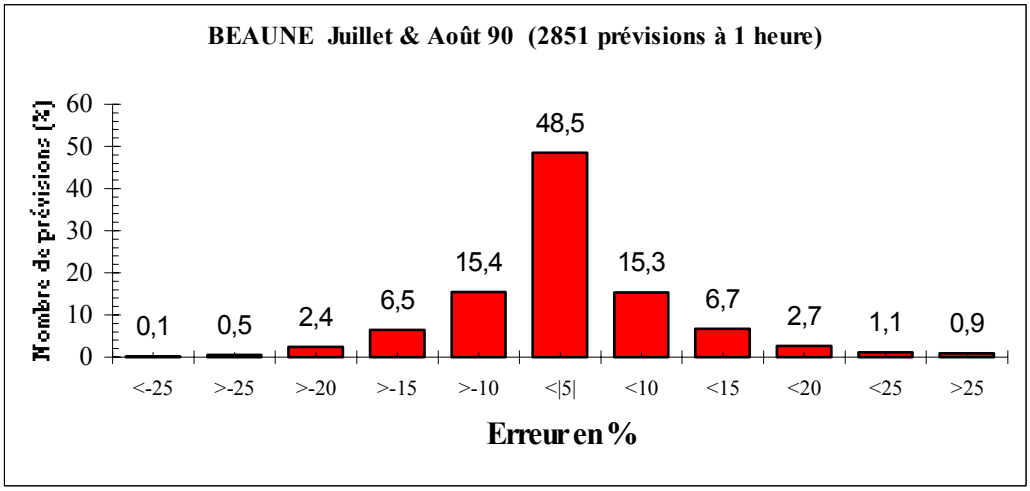
Tandis que les deux premières étapes se réalisent en temps différé, la troisième est destinée au temps réel

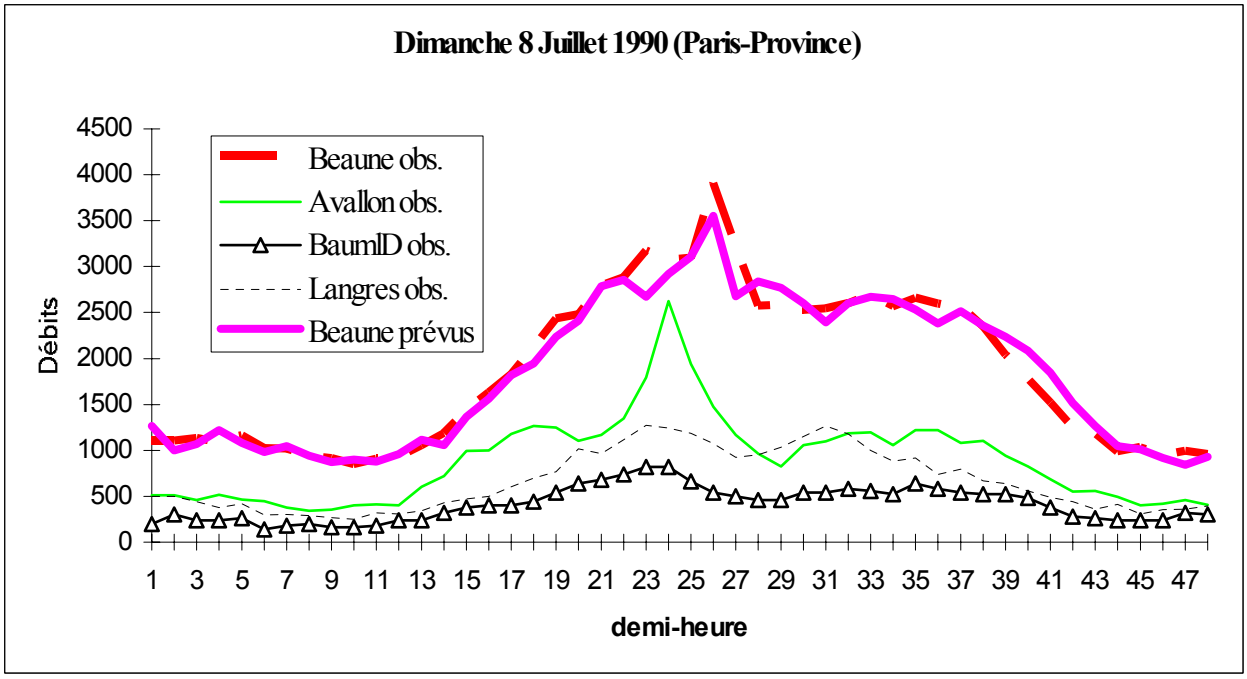
RESULTATS

Depuis 1990 ce modèle est devenu un outil de prévision au PC la SAPRR à Beaune ; de plus, il a été utilisé dans le cadre du programme européen DRIVE I, pour les projets MONICA et SOCRATES (Göteborg) et il fait partie intégrante du module autoroutier conçu dans le cadre du programme européen DRIVE II (Grand Lyon).

Ici, nous nous contentons de présenter les résultats du test de ce modèle éffectué avec l'aide de la SAPRR au triangle de Beaune pendant l'été 1990. Le point de prévision est situé à Beaune (Pk 305,3) dans le sens Paris-Province, après le confluent de trois autoroutes A6 (en provenance de Paris via Avallon), A36 (en provenance de Mulhouse via Baume-les-Dames) et A31 (en provenance de Nancy via Langres). Les données utilisées pour le calibrage, concernant les mois de Juillet et Août de 1984 à 1989, proviennent de quatre stations SNRD.

Dans ce test, on a effectué 2851 prédictions dont 1412 en Juillet et 1439 en Août. Dans l'ensemble, 2.357.429 véhicules ont été détectés par le capteur situé à Beaune (débit moyen de 1.654 véhicule heure). Le modèle ATHENA à l'horizon d'une heure a prévu 2.360.751 véhicules (très légère surestimation). La répartition des erreurs de prévision à une heure fait apparaître que dans 79% des cas la valeur absolue de cette erreur est inférieure à 10%. Les schémas ci-dessous représentent l'histogramme des erreurs et l'exemple d'une journée entière de prédiction.





CONCLUSION

Le bilan global des divers tests et applications confirment que l'objectif initial de ce modèle (rester en deçà d'un seuil de 10% d'erreur) est atteint. Par ailleurs, ces applications montrent une certaine faiblesse du modèle à prédire les très grandes irrégularités dues en général aux congestions. Pour combler ce genre de lacunes, nous étudions actuellement, d'autres approches qui font appel à des modélisations non linéaires utilisées dans l'étude des systèmes dynamiques.

COPYRIGHT : Reproduction autorisée sous réserve d'en mentionner l'origine
PRIX :150.00 Francs T.T.C. (T.V.A. 2.10 %). Les commandes doivent être adressées à I'INRETS, Services publications - 2. av. du Général Malleret-Joinville - 94114 ARCUEIL CEDEX, et être accompagnées d'un chèque bancaire ou postal à l'ordre de : I'Agent Comptable de I'INRETS. Déduire la TVA pour l'étranger hors CEE, Pour les pays de la CEE se reporter à la réglementation intracommunautaire en matière de TVA.

SUMMARY

ATHENA

A METHOD FOR SHORT TERM INTER-URBAN TRAFFIC FORECASTING

ATHENA's main objective is to forecast, on a short horizon (half-an-hour to two hours) the traffic of a road section.

Any mathematical modelling requires a certain number of assumptions. ATHENA's first assumption is that the traffic of the near future (half-an-hour to two hours) depends on the traffic of the moment and also of the near past. The second assumption is that this dependence can be defined by a linear mode1 formed on the basis of historical data. This linear assumption is attenuated by the forming of a series of models, each mode1 representing a certain type of traffic.

The method comprises three main steps : the typological analysis of historical data, modelling and finally the very forecasting.

The global balance of the applications confirms that the initial objective of this mode1 (remain under a threshold of 10%. errors) is reached.

RESUME

ATHENA

PREVISION A COURT TERME DU TRAFIC SUR UNE SECTION DE ROUTE

L'objectif principal d' ATHENA est de prédire, pour un horizon court (une demi-heure à deux heures), le trafic d'une section de route. Toute modélisation mathématique impose un certain nombre d'hypothèses.

La première hypothèse d'ATHENA est que le trafic de l'avenir proche (une demi-heure à deux heures) dépend du trafic du moment et aussi du passé proche, trafics que l'on peut observer au point même de la prévision mais aussi en un certain nombre de points amont. La deuxième est que cette dépendance peut se définir par un modèle linéaire constitué à partir des données historiques. Cette hypothèse de linéarité est atténuée par la constitution d'une série de modèles, chaque modèle traduisant un certain type de trafic,

La méthode comprend trois étapes principales: l'analyse typologique des données historiques, la modélisation et la prédiction.

Le bilan global des divers tests et applications confirment que l'objectif initial de ce modèle (rester en deçà d'un seuil de 10% d'erreur) est atteint.

TABLE DES MATIERES

I. Introduction 4
I.1 La prédiction du trafic, un besoin des exploitants 4
I.2 Les moyens de mesures 4
I.3 Le traitement des mesures 5
I.3.1. Les considérations temporelles des mesures 5
I.3.2. Les considérations spatiales des mesures 5

II CONCEPTION DU MODELE ATHENA 6
II.1. Première étape : l'analyse typologique 7
II.1.1. Les données historiques 7
II.l.2 L'agrégation des débits 8
II.l.3 La Constitution des profils 8
II.l.4 La classification des profils 10
II.l.5 La validité de la classification 10
II.1.6 La classification mathématique comparée à la classification calendaire 11
II.2 Deuxième étape : la modélisation 11
II.2.l La formulation du modèle 11
II.2.2 Validité du modèle de régression 13
II.3 Troisième étape : la prévision 13
11.3.1 Le choix de classe 13
II.3.2 L'application du modèle de régression 14
II.4. Les critères d'évaluation des modèles ATHENA 14
REFERENCES 17

III. UN EXEMPLE D'APPLICATION 19
III.1 Les données de calibrage 19
III.2 La constitution des profils et leur classification 22
III.3 Le croisement entre la classification mathématique et la classification calendaire 24
III.4 Le calibrage des modèles 27
III.5 Le résultat de l'application d'ATHENA (modèle n°3) en Juillet et Août 90 28
III.6 Explication des erreurs fortes (>15%) 30
III.7 Erreurs dues aux fluctuations importantes de débits 31
III.8 Un exemple d'une journée de prévision 31

IV CONCLUSIONS 34
IV.1 L'originalité du modèle 34
IV.2 Les points forts 34
IV.3 Les points faibles 34

ANNEXE A CLASSIFICATION (rappel de quelques notions) 36
ANNEXE B REGRESSION MULTIPLE (rappel de quelques notions) 38
ANNEXE C MODE D4EMPLOI 41

I. Introduction

I.1 La prédiction du trafic, un besoin des exploitants

Les exploitants disposent de données de trafic de plus en plus nombreuses et recueillies à grand frais; cependant il n'existe pas suffisamment de logiciels qui les aident à tirer parti de tout ce que ces données signifient.

Notamment, bien qu'il existe des prévisions du trafic à long terme (Bison Futé quelques mois à l'avance) et à moyen terme (prévisions régionales 1 semaine à l'avance), les exploitants ne disposent pas de prévisions à plus court terme. Pourtant celles-ci constituent un indicateur important pour la régulation ; elles permettent notamment de prendre, aussi bien que de rapporter, plus rapidement les mesures de régulation, et d'éviter ainsi certaines congestions. L'exploitant a aussi la possibilité de tenir informés les usagers de ces prévisions. La méthode proposée n'a pas la prétention de tenir compte des effets induits par ce type d'informations.

I.2 Les moyens de mesures

Les moyens d'observation (capteurs routiers) existent et se développent depuis plusieurs années; on peut citer :

Le Système National de Recueil des Données (SNRD) composé de 81 stations fournissent les débits et les vitesses par demi-heure.

le système "REGA" de la Société des Autoroutes Paris-Rhin-Rhône, qui comprend depuis 1988 28 capteurs entre Paris et Lyon (débits 6 minutes, vitesse, taux d'occupation, pourcentage de véhicules longs)

le système "MISTRAL" de la société des Autoroutes du Sud de la France (50 capteurs 6 minutes).

D'autre part le Ministère de l'équipement pour harmoniser les échanges des données routières a fait développer le standard SIREDO, qui intègre aussi bien les mesures individuelles et instantanées (vitesse, intervalle, longueur,..) que les mesures séquencées sur 1 minute, 6 minutes ou un jour.

I.3 Le traitement des mesures

Tirer parti des données, cela veut dire, par exemple que, lorsqu'une pointe de trafic est observée en un point du réseau, l'objectif est de prédire à quel moment et avec quelle amplitude cette pointe va se présenter en un point situé en aval ; ou d'analyser les variations horaires du trafic pour en dégager les régularités : celles-ci seront utilisées dans la prévision,
Une grande partie de la variabilité des données peut être expliquée par des considérations temporelles et/ou spatiales.

I.3.1. Les considérations temporelles des mesures

Le trafic en un point est fonction de l'heure et du jour. Une méthode (utilisée par certains exploitants) consiste à chercher dans le passé une référence valable (jour de même nature, même heure) et à en déduire le trafic des prochaines heures. Notre approche n'est pas très différente sur le fond : la référence n'est toutefois pas constituée d'un seul jour, dont l'évolution peut être exceptionnelle, mais d'une classe de jours semblables.

I.3.2. Les considérations spatiales des mesures

Une connaissance complète de ce qui se passe à un moment donné dans le réseau permet théoriquement une prévision exacte de ce qui va se passer un peu plus tard en aval; cela suppose que l'on connaisse les destinations des usagers, que leur vitesse soit constante (ou modélisable par une "loi" de trafic). La connaissance du trafic étant incomplète, la corrélation que l'on peut calculer entre un point aval et les points amont ne sera jamais égale à 100 %, mais sera cependant un élément significatif pour la prévision.
Un certain nombre de logiciels de prévision du trafic ont déjà été développés :

- Un filtre adaptatif a été utilisé (cf [3]) pour prévoir une heure à l'avance le trafic arrivant à Beaune (au confluent des autoroutes A6, A31, A36) : on s'est servi des comptages demi-heure de 2 points situés à 100 Km en amont de Beaune (sur A6 et

A31, les données de l'autoroute A36 n'étant pas disponibles) ; le filtre revenait à sommer ces données avec des coefficients de pondération "adaptatifs" : l'erreur de prévision une fois constatée servait à modifier ces coefficients.

- Un modèle de régression a été testé (cf. [3]) pour prévoir le trafic sur l'autoroute du Sud, le trafic prévu à Tain (au Nord de Valence) à h+1 heures dans le sens Sud-Nord était expliqué par le dernier trafic connu (à h) à Orange, (Orange est approximativement à une heure de conduite de Tain; l'erreur quadratique moyenne (cf. II.4), atteignait 13 % pour un jour de grands retours.
- Dans le cas des grands départs, un site immédiatement en aval de Lyon avait été choisi : là, la corrélation amont-aval n'était plus significative du fait de la présence entre l'amont et l'aval d'une ville générant un trafic important. Le trafic était extrapolé de façon purement statistique; mais la précision des prévisions (16%) s'est révélée insuffisante ; ceci provenait en partie du fait que le calibrage du modèle avait été effectué (heure par heure) sur un ensemble de données (les mois de Juillet et Août 1985) peu homogène : certaines de ces journées correspondent à des grands départs, d'autres non : la droite des moindres carrés passait par une moyenne peu significative, les prévisions étaient "plates", les coefficients du modèle étaient voisins de 0 ou de 1.

II CONCEPTION DU MODELE ATHENA

L'objectif principal d' ATHENA est de prédire, pour un horizon court (une demi-heure à deux heures), le trafic d'une section de route. Ce type de prédiction est plutôt destiné aux gestionnaires du trafic routier qu'aux usagers de la route. Cette finalité exige, pour qu'il soit utilisable, une certaine précision dans la prédiction, dépassant la faculté de prédiction des gestionnaires. Toutefois nous délimitons notre objectif en faisant la distinction entre la prévision de la demande et celle de l'écoulement du trafic. La vraie cible d'ATHENA est la prévision de la demande et non pas celle de l'écoulement du trafic qui ne peut se faire que par un modèle de simulation.

Toute modélisation mathématique impose un certain nombre d'hypothèses. La première hypothèse d'ATHENA est que le trafic de l'avenir proche (une demi-heure à deux

heures) dépend du trafic du moment et aussi du passé proche. La deuxième est que cette dépendance peut se définir par un modèle linéaire constitué à partir des données historiques. Cette hypothèse de linéarité est atténuée par la constitution d'une série de modèles, chaque modèle traduisant un certain type de trafic. Ainsi l'hypothèse de linéarité de base se réalise dans un processus non linéaire ou linéaire par morceau.

La méthode comprend trois étapes principales : l'analyse typologique des données historiques, la modélisation, enfin la prévision proprement dite (cf. page 16 l'algorithme d' ATHENA) .

II.1. Première étape : l'analyse typologique

II.1.1. Les données historiques

En général le trafic d'une section de route est décrit par un certain nombre de variables issues d'un capteur. Quelles variables doit-on prendre pour décrire le trafic d'un jour ou d'une période donnée? Les taux d'occupation (exprimés en pourcentages) ne comportent que deux chiffres significatifs et sont donc peu précis; les vitesses évoluent peu en général et se prêtent donc mal à l'analyse, et les débits qui peuvent être horaires, demi-horaires ou relatifs à d'autres fractions de l'heure. Tandis que les deux premières variables expriment l'écoulement du trafic, le débit, lui, est un très bon indicateur de la demande et correspond au mieux à l'objectif que nous nous sommes fixés.

Les données historiques nécessaires à l'utilisation d'ATHENA sont donc, uniquement, des mesures de débits. En ce qui concerne la durée de I'historique, elle doit être ni trop courte, pour qu'elle ait un sens statistique, ni trop longue, pour tenir compte des dernières évolutions physiques (le nombre de voies, leur largeur, l'existence de routes alternatives....) de la section de route en question. Ceci est encore plus important quand on veut aussi faire intervenir les effets du trafic des points situés à l'amont de cette section (cf. infra). A titre d'exemple si l'on s'intéresse au trafic pendant le mois d'Août sur une section de route en rase campagne, il suffit que la base de données couvre de cinq à six mois d'Août successifs qui constituent la durée de I'historique. Cette base de données a la forme d'un tableau tel que chaque ligne, identifiée par la date d'un jour, représente les mesures des débits

en ce jour. Le nombre de lignes de ce tableau est égal au nombre de jours de l'historique. Le point de prévision et les éventuels amonts constituent chacun un tableau séparé.

II.I.2 L'agrégation des débits

La périodicité d'acquisition des débits dépend, bien sûr, du type d'équipement en capteurs; selon l'équipement il peut être, par exemple, 2, 10 ou 90 mesures par heure (i.e. respectivement, une mesure chaque demi-heure, toutes les 6 minutes ou toutes les 40 secondes). Mais la périodicité d'agrégation des débits est un choix qui varie selon la nature du trafic (urbain , suburbain ou autoroutier). Elle ne doit pas être trop petite, parce que sur une trop courte période, les débits étant instables, ils se prêtent mal à l'analyse. Elle ne doit pas être trop longue parce qu'elle entraîne un certain lissage qui peut masquer certains phénomènes périodiques (e.g. période de pointe).

II.I.3 La Constitution des profils

Avant d'aborder l'analyse typologique, fixons quelques notations qui faciliteront la démarche.

Les notations

θ la périodicité d'agrégation des débits (e.g. 6 minutes, ou une demi-heure),

h l'horizon de la prédiction (exprimé en unité de temps θ)

t l'instant de la journée où on effectue la prédiction (exprimé en θ)

n_l l'intervalle du temps avant l'instant de la prévision t (exprimé en θ)

$Q[j,t]$ le vecteur de débits à n composantes relatif au jour j à l'instant t.

$$Q[j,t] = [q_{ji}^t \,|\, i = t - n_1, t + h]$$

Où q_{ji}^t est le débit au jour j à l'instant i

$$n = n_1 + h$$

$P[j,t]$ le vecteur profil à n composantes relatives au jour j à l'instant t:

$P[j,t] = [P_{ji}^t \,|\, i = t - n_1, t + h]$

où $P_{ji}^t = q_{ji}^t / [\sum_m q_{jm}^t \,|\, m = t - n_1, t + h]$

On sait que le trafic d'un jour dépend essentiellement de deux sortes de variations : la variation journalière (opposant les jours normaux aux jours de grands départs), et la variation horaire (opposant les heures creuses aux heures de pointe). Notre objectif étant de faire une prédiction de débit à l'instant t pour l'horizon h, nous devons, donc, nous intéresser en premier lieu à la variation liée à la période correspondante. Pour l'analyse typologique nous raisonnons sur les profils de débits $P[j,t]$. Les variations journalières disparaissent alors, ce qui d'une part diminue la variance du phénomène (sans effet négatif sur les prévisions comme on le verra par la suite) et qui d'autre part permet d'effacer l'augmentation du trafic (i.e. effet de taille) d'une année sur l'autre : ceci est positif, toutes les années de l'historique seront utiles aux prévisions.

Le choix de n_1 (i.e. l'intervalle du temps avant l'instant t de la prévision) dépend de l'unité θ et de la nature du trafic (urbain, suburbain ou autoroutier) ; elle ne peut être fixée que cas par cas et nécessite une connaissance parfaite de l'ensemble des données historiques dont on dispose. Par ailleurs, si les données sont suffisamment importantes, on doit procéder, au préalable, à un regroupement saisonnier du trafic (e.g. été, mi-saison, hiver...). La classification sera effectuée, séparément, sur chacun des groupes de mois. On remarque ainsi que dans la méthode proposée, ATHENA, on ne fait intervenir aucune information a priori (e.g. jour férié, heure chargée...) et nous nous contentons uniquement des informations sous-jacentes aux données disponibles.

L'introduction, dans le profil de l'instant t, des débits de l'intervalle de temps $[t, t+h[$ permet d'associer leur variations à celles des débits de l'intervalle $[t - n_1, t]$ Dans l'étape de la prévision (en temps réel) nous ne disposons pas des débits de l'intervalle $[t, t+h[$; comme on le verra plus loin, nous sommes obligés de passer par un calcul d'optimisation pour reconstituer ce profil, c'est pourquoi h ne doit pas être trop grand par rapport à n_1.

II.I.4 La classification des profils

L'objet de la classification est de répartir les données historiques en un certain nombre de classes. A l'intérieur d'une classe la variance des phénomènes étudiés est plus faible, ce qui favorise l'extrapolation des phénomènes. On peut même espérer que les phénomènes différents se répartissent dans des classes différentes. Le mécanisme d'un algorithme de classification est développé en annexe A; on rappelle, simplement ici, que dans toute méthode de classification il y a trois notions à définir : les objets à classer, la distance entre les objets et le critère d'agrégation d'un objet à une classe. Les objets à classer sont les profils $P[j,t]$; dans l'algorithme choisi (centre mobile) la distance entre les objets est la distance euclidienne, et le critère d'agrégation est la distance euclidienne entre un objet et le centre de gravité d'une classe.

Chaque classe c d'effectif e_c, sera représentée par son "profil moyen" le vecteur $P[c,t]$ à n composantes :

$$P[c,t] = [\, p^t_{ci} \,|\, i = t - n_1, t + h]$$

Où $p^t_{ci} = [\sum_j p^t_{ji} \,|\, j \in c] / e_c$,

Dans l'étape de la prévision (cf infra), ces profils moyens serviront au choix de classe.

II.I.5 La validité de la classification

Les résultats de la classification ne sont pas automatiquement acceptés tels quels. Par exemple si une classe ne comporte qu'un seul élément, on peut suspecter une erreur de données ou un événement exceptionnel, qu'il faut corriger, et ensuite recommencer la classification. Un effectif minimum par classe est nécessaire pour que la modélisation ultérieure de chaque classe soit possible. Ceci implique, comme la taille de l'échantillon est limitée, certaines contraintes sur le nombre de classes.

Un autre moyen de valider la classification est de mesurer l'information qu'elle apporte (cf. annexe A). En général ce type d'information est traduit par le rapport entre la variance entre les classes sur la variance totale; l'algorithme de classification choisi (centre mobile), optimise cette information. Une valeur trop faible de ce rapport sera source de doute quant à la validité de la classification.

II.1.6 La classification mathématique comparée à la classification calendaire

Une classification calendaire ne peut séparer l'ensemble des données historiques, qu'en un certain nombre de types du trafic comme les jours de semaines, les week-end, et les fêtes. Dans la classification mathématique (i.e. automatique) telle qu'elle est appliquée dans ATHENA, ce qui importe, c'est la variation du trafic à l'intérieur d'un intervalle de temps qui

est une partie de la journée. Dans les classes déterminées par cette méthode, on retrouve, pour une part, les types calendaires et pour une autre part, des ressemblances entre jours qui ne sont pas évidentes du point de vue calendaire, mais qui sont dégagées par les classes (cf. infra). A titre d'exemple si la périodicité d'agrégation des données est la demi-heure (i.e. $\theta = 30$ minutes), pour une valeur fixée de n_1, pour un horizon de prévision donné et pour chaque demi-heure de la journée nous procédons à une classification. Si l'on suppose que chaque classification produit 4 classes, pour les 48 demi-heures de la journée nous aurons 192 classes. Dans certains types du trafic (e.g. autoroutier), il n'est pas nécessaire que la classification soit aussi répétitive que θ; dans cet exemple la classification faite toutes les heures pourrait suffire et le nombre de classes serait diminué de moitié (i.e. 96).

II.2 Deuxième étape : la modélisation

II.2.I La formulation du modèle

Les variables explicatives introduites sont déduites des considérations temporelles et spatiales. Les considérations temporelles permettent de tenir compte de l'évolution dans le temps du trafic de la section de route concernée (i.e. le point même de prévision), les considérations spatiales font intervenir l'effet d'écoulement du trafic entre une section située à l'amont et le point de prévision. Soient :

$A = [\alpha_0 \quad \alpha_1 \quad \alpha_2 .....\alpha_\lambda]$ les coefficients relatifs au point de prévision ,

$Q_0 = [q_t^0 \quad q_{t-\theta}^0 \quad q_{t-2\theta}^0 .....q_{t-\lambda\theta}^0]$ les débits relatifs au point de prévision,

$B = [\beta_0 \quad \beta_1 \quad \beta_2 .....]$ les coefficients relatifs au premier amont et

$Q_1 = [q^1_{t-\delta} \quad q^1_{t-\delta-\theta} \quad q^1_{t-\delta-2\theta} .....]$ les débits relatifs au premier amont.

Où q^0_t est le débit, à l'instant t, du point de prévision et $q^1_{t-\delta}$ est le débit, à l'instant $t-\delta$, du premier amont; δ (exprimé en θ) est le temps de parcours moyen entre cet amont et le point de prévision diminué de l'horizon h. Ainsi un véhicule observé en amont à t-δ passera au lieu de prévision à t+h.

Pour chaque classe issue de l'étape précédente on établit un modèle linéaire (régression multiple) reliant le débit de l'instant t+h aux débits des instants t, $t-\theta$, $t-2\theta$, ……..$t-\lambda\theta$ ainsi qu'aux débits de points situés en amont immédiat (λ est un entier).

$$q_{t+h} = Q_0 A' + Q_1 B' + ... \quad (1)$$

Où A' et B' sont, respectivement, des transposés des vecteurs A et B. Les nombres de composantes des vecteurs Q_0 , Q_1, A, et B, sont limités à $\lambda + 1$. La partie droite de la relation (1) peut s'étendre, si nécessaire à plusieurs amonts.

Nous n'avons pas gardé les profils au stade de la régression parce que le besoin est bien de prévoir des débits. Cependant, comme les classes ne sont homogènes qu'en profils, et que les débits peuvent y être très dispersés, il ne faut pas centrer les débits autour d'un débit moyen sans signification, il faut que la droite de régression passe par l'origine, (sans terme constant).

Ce modèle a été doublé par un modèle où les coefficients de régression ont été contraints à être positifs (cf. annexe B); un coefficient négatif indique souvent un excès de variables explicatives. Lorsqu'il traduit l'influence d'un point amont, un coefficient positif signifie la part du trafic amont qui passe au point de prévision.

Le nombre de coefficients à retenir, dépend évidemment de l'effectif de la classe. La multiplication des coefficients de régression, améliore certes l'ajustement de la droite de régression, mais dans le cas où les variances des coefficients sont fortes, elle détruit la qualité de la précision; on verra en annexe B, que l'intervalle de confiance de la prévision est lié à ces variances.

II.2.2 Validité du modèle de régression

On utilise deux critères, l'un relatif et l'autre absolu, pour mesurer la validité de ce modèle. Le premier est la part de la variance expliquée par la droite de régression (i.e. coefficient de détermination R^2). Le deuxième critère est I'estimateur de l'écart type des résidus (cf. annexe B), sa valeur est comparée avec le débit moyen pour la période θ. De plus, sous certaine hypothèse de normalité, on peut calculer un intervalle de confiance pour chacun des coefficients.

II.3 Troisième étape : la prévision

Cette étape qui s'effectue en temps réel se réfère pour chaque instant t à la classification et à la modélisation de ce même instant. Elle comporte deux parties : le choix de la classe des débits, puis l'application du modèle de régression.

11.3.1 Le choix de classe

Pour choisir parmi les classes de l'instant t, la classe la plus proche, on doit calculer, à l'aide des derniers débits disponibles, un profil de débits ayant le même nombre de composantes que les profils moyens des classes (i.e. $n = n_1 + 1 + h$). Or on sait qu'en temps réel, à l'instant t, les h mesures de débits à venir ne sont pas encore connues. Le choix de classe sera donc fait en passant par un calcul d'optimisation. Soient :

q_t débit à l'instant t,

x_c^t l'inverse du débit total (inconnu) dans l'intervalle du temps $[t - n_1, t + h]$,

$P[c,t]$ le profil moyen de la classe c de l'instant t. On établit pour chaque classe c le critère (F_c) :

$$F_c = [\sum_i (q_i x_c^t - p_{ci}^t)^2 \,|\, i = t - n_1, t]$$

La valeur optimale de x_c^t s'obtient en dérivant cette fonction. La distance entre le profil de l'instant t et le profil moyen de la classe c est calculée par :

$$d_c^2 = [\sum_i (q_i x_c^t - p_{ci}^t)^2 \,|\, i = t - n_1, t]$$

Le minimum de celle-ci, calculé pour toutes les classes, fait choisir la classe dont le trafic est

le plus semblable au trafic écoulé entre $t - n_1$ et t. L'hypothèse implicite est que le trafic, qui va survenir durant l'intervalle $[t+1\ ,t+h]$ subira la même variation en profil (non pas en volume) que le trafic du même intervalle de la classe choisie. La précision de cette optimisation dépend sensiblement de la longueur de l'horizon h.

II.3.2 L'application du modèle de régression

Cette partie, consiste à faire l'usage du modèle de régression de la classe choisie. Pour chaque valeur prédite on peut, à l'aide de la matrice de covariances des variables explicatives et l'écart type des résidus, calculer un intervalle de confiance (cf. annexe B).

Dans le cas où l'on dispose des derniers changements de capacité de la section de route considérée, la prévision est écrêtée en conséquence.

II.4. Les critères d'évaluation des modèles ATHENA

Nous avons considéré, deux critères d'erreur, un global E_g, tenant compte de l'ensemble des prévisions pour un point, effectuées pendant une période (e.g. un mois) et l'autre partiel

E_p^t, calculé pour chaque prévision. Soient :

$q_t =$ le débit observé à l'instant t,

alors

$\Delta_t =$ l'écart entre le débit prévu et le débit observé,

$E_g^2 = [\sum_t \Delta_t^2 / \sum_t q_t^2 \,|\, t \in$ péride des prévisions$]$

$E_p^t = |\Delta_t| / q_t$.

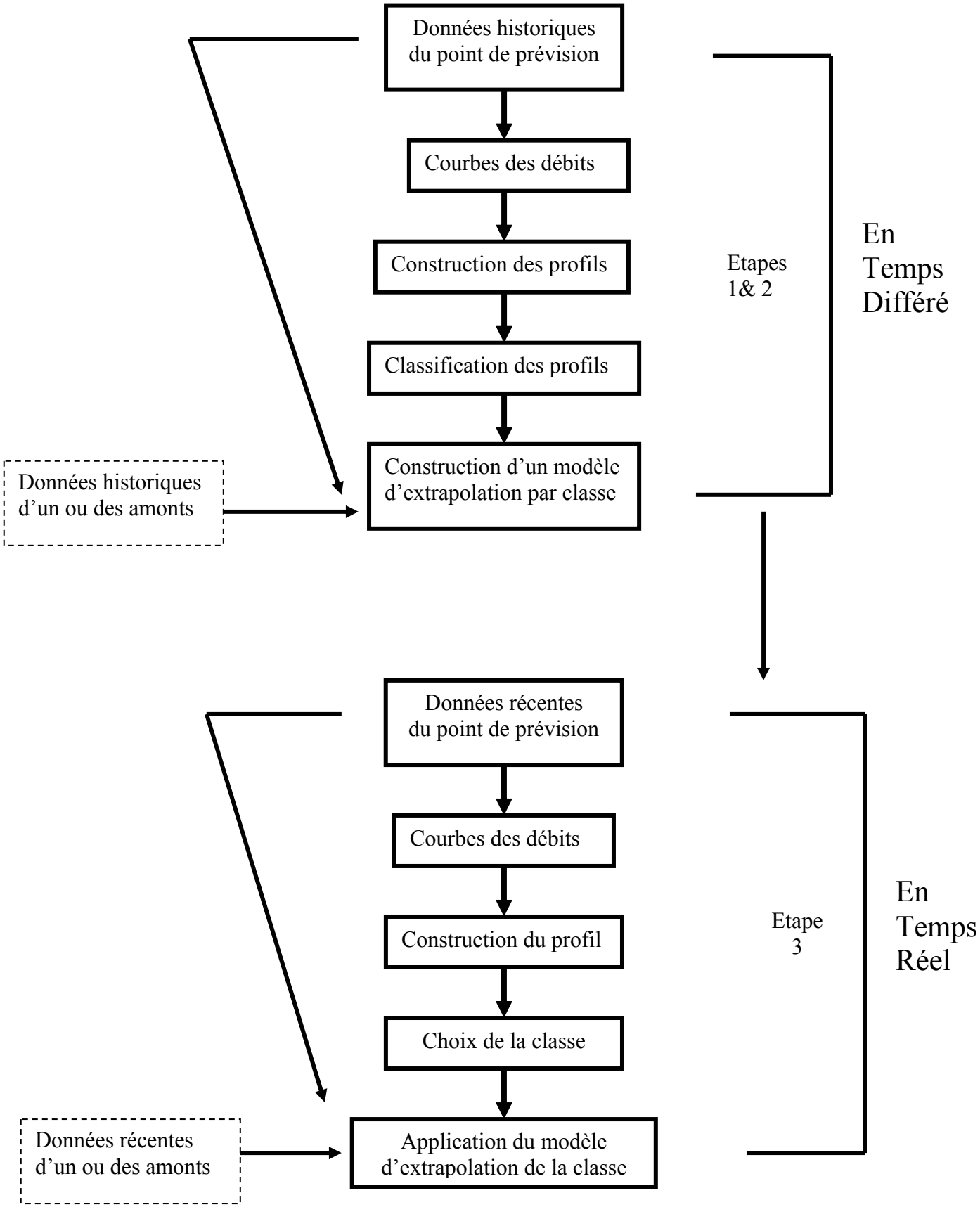
Le critère global sert à fournir une idée globale de la précision du modèle. Le critère partiel donne une idée de la fiabilité ou de la robustesse du modèle ; il permet de dire, si l'on se fixe un seuil en pourcentage d'erreur partielle, combien de prévisions ont été bonnes. Sa distribution selon le niveau des débits réels est un bon moyen pour analyser la variation de ce critère d'erreur selon l'importance des débits. Une erreur forte (en pourcentage) sur un débit faible est, peut être, moins importante qu'une erreur faible (en pourcentage) sur un débit fort. D'autre part si le modèle est utilisé sur une période continue (e.g. un mois ou plus), la

distribution de E_p^t, selon l'heure de la journée, nous permet de localiser les éventuelles erreurs systématiques.

Remarque: Un autre critère global d'erreur, exprimé en nombre de véhicules, "RMSE" est calculé de la manière suivante :

$RMSE^2 = [\sum_t \Delta_t^2 / n | t \in$ période des prévisions$]$ n = nombre de prévisions

Algorithme du modèle ATHENA


Données historiques du point de prévision
Courbes des débits
Construction des profils
Classification des profils
Construction d'un modèle d'extrapolation par classe
Données historiques d'un ou des amonts
Etapes 1& 2
En Temps Différé
Données récentes du point de prévision
Courbes des débits
Construction du profil
Choix de la classe
Application du modèle d'extrapolation de la classe
Données récentes d'un ou des amonts
Etape 3
En Temps Réel

REFERENCES

1. ANDERBERG M. R. Cluster Analysis for applications ACADEMIC PRESS 1973
2. AFIFI A. A., AZEN S.P. Statistical Analysis a Computer Approche ACADEMIC PRESS 1972
3. ARON M., DIEP D., DAVEE MM. Prévision de trafic- Application à l'été 1986 (INRETS 1987).
4. ARON M., DANECH-PAJOUH M., STOVALL T., Une nouvelle méthode de prévision du trafic autoroutier à court terme Revue générale des routes et aérodromes No 674 mai 1990
5. CAILLIEZ F., PAGES J. P. Introduction à l'analyse des données SMASH 1976
6. DANECH-PAJOUH M., ARON M. ATHENA Une Méthode de Prévision à Court terme du Trafic Autoroutier RTS n° 28 Décembre 1990
7. DANECH-PAJOUH M., ARON M. ATHENA Application à Beaune en Juillet et Août 1990 Rapport convention INRETS-SAPRR Septembre 1990
8. DANECH-PAJOUH M., ARON M. ATHENA Mode d'emploi version Juin 1991, Rapport interne INRETS
9. DIEP D. Une méthode de prévision du trafic à court terme, Application à l'opération PaIomar 1985 INRETS (Octobre 1985)
10. FLORENS J.P., IVALDI M. Estimation des taux horaires de trafic en différents points du réseau d'ASF en fonction du taux moyen journalier annuel GREMAC Université de Toulouse Février 1989
11. DRIVE 1056 "MONICA" WP E1.1 Deliverable n° 17 Prepared by INRETS and LCPC
12. LEMAITRE G. Routes et villes: vers une harmonisation de la mesure de trafic, RTS n° 33 Mars 1992
13. MADRE J.L., LAMBERT T. Prévisions à long terme du trafic automobile CREDOC Mai 1989
14. MODULAD version 2.1 INRIA

15. FOUILHE P. Bison Futé, Bilan technique sur 12 ans d'activités (1976-1987) DSCR Décembre 1987
16. ROUX M. Algorithmes de classification MASSON 1985
17. SAPORTA G. Probabilité analyse des Données et Statistique TECNIP 1990
18. TETREL O., DANECH-PAJOUH M., Relation entre débit horaire et taux d'accidents ONSER Août 1974
19. WATSON SM An intoduction to adaptive forecasting techniques for on-line traffic flow estimation ITS December 1992

III. UN EXEMPLE D'APPLICATION

Depuis 1990 ce modèle est devenu un outil de prévision au PC de Beaune de la SAPRR; de plus il a été utilise dans le cadre du programme européen DRIVE I, pour les projets MONICA et SOCRATES (Göteborg) et il fait partie intégrante du module autoroutier conçu dans le cadre du programme européen DRIVE II du projet LLAMD (Lyon). Avant d'être sous sa forme opérationnelle d'aujourd'hui, il a été testé à plusieurs reprises sur les sites suivants:

à Nîmes, à Bollène et à Villefranche en été 1989 et 90 avec l'aide du SETRA; à Cevin et à Moutiers en hiver 1990 avec l'aide de la DSCR

à Vienne et à Montélimar en été 1991 avec l'aide d'ASF;

à Beaune en été 1990 avec l'aide de la SAPRR.

Dans ce chapitre nous nous contentons de développer ce dernier test, comme un exemple d'application.

III.1 Les données de calibrage

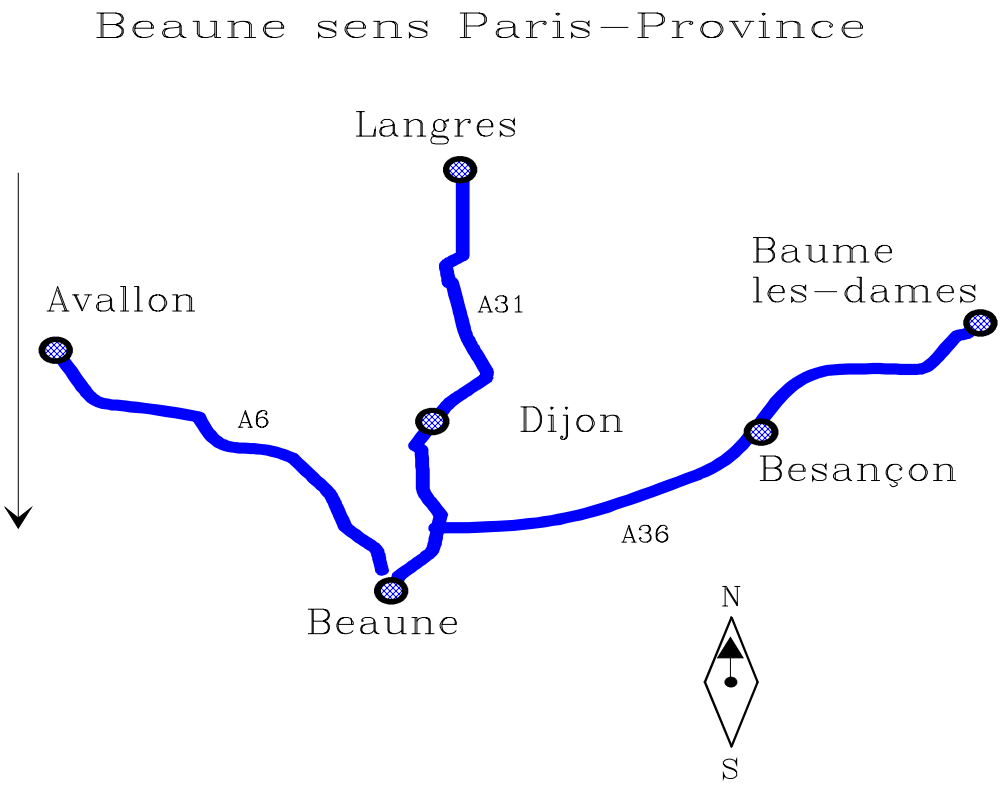
Le point de prévision est situé à Beaune (Pk 305,3) dans le sens Paris-Province,
après le confluent de trois autoroutes (cf. plan ci-après) A6 (en provenance de Paris via Avallon), A36 (en provenance de Mulhouse via Baume-les-Dames) et A31 (en provenance de Nancy via Langres). Dans cette application on considère que les trafics des points situés
en amont (Avallon, Baume-les-Dames, Langres) influent sur le trafic du point de prévision.
Les données utilisées pour le calibrage proviennent de quatre stations SNRD, concernant les mois de Juillet et Août de 1984 à 1989. La périodicité de ces données est de 30 minutes; nous avons conserve la même périodicité pour l'ensemble des calculs de calibrage, sauf pour la partie classification où nous avons agrégé les données sur 60 minutes (cf. infra.). La procédure de calibrage suppose l'épuration des données en cas d'anomalie due aux capteurs ou au trafic (e.g. le trafic est nul pendant plusieurs périodes d'acquisition). Le tableau ci-

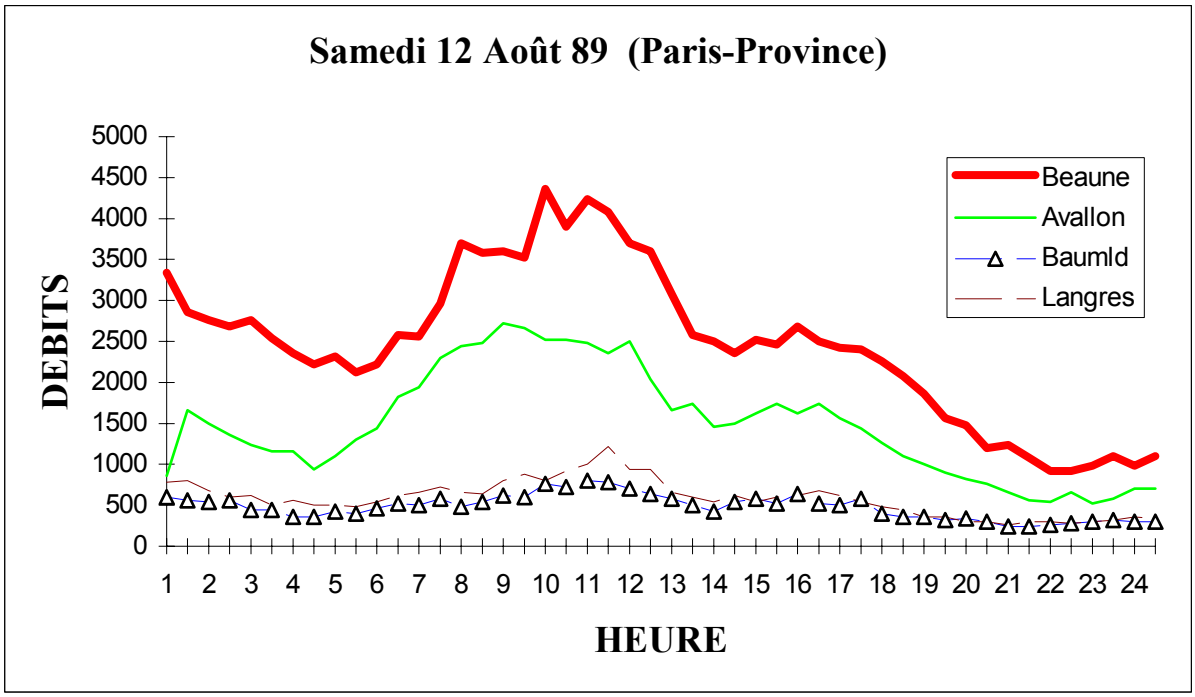
dessous indique pour chaque capteur le nombre de jours valables.

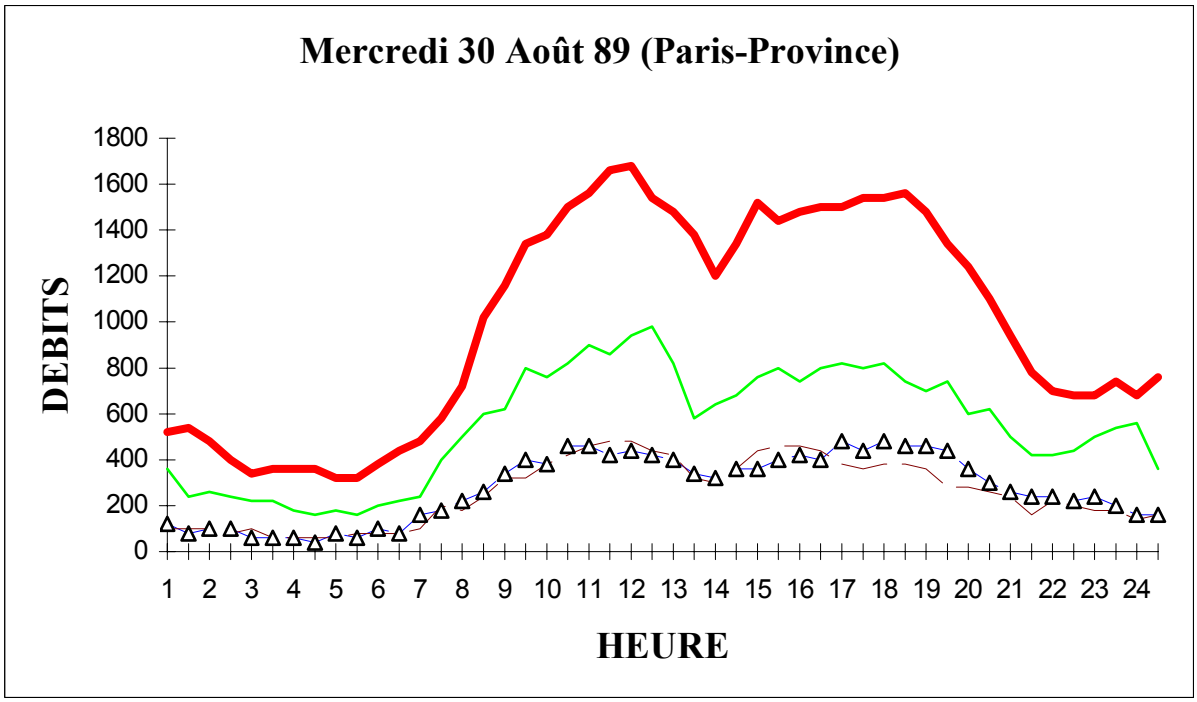
Compteurs SNRD (sens Paris-Province)

Point	Autoroute	Pk	Nb. De jours
Beaune	A6	305,3	347
Avallon	A6	221,8	317
Baume-les-Dames	A36	91,0	285
Langres	A31	98,8	285

Pour mieux saisir l'influence du trafic des points amont, nous avons trace pour deux jours a priori complètement différents les courbes des débits des quatre capteurs correspondants (voir graphiques ci-dessous). On constate ainsi que la courbe de Beaune suit bien, avec un décalage dû au temps de parcours, les mouvements de la courbe d'Avallon ; les deux courbes de Langres et de Baume-les-Dames ont un comportement parallèle et presque avec le même volume du trafic.




Samedi 12 Août 89 (Paris-Province)
DEBITS
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
HEURE
Beaune
Avallon
Baumld
Langres


Mercredi 30 Août 89 (Paris-Province)
DEBITS
1800
1600
1400
1200
1000
800
600
400
200
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
HEURE

III.2 La constitution des profils et leur classification

En respectant les notations du paragraphe 1.2.3, nous précisons ici les valeurs affectées à différents paramètres. La périodicité d'agrégation des débits θ est fixée à 30 minutes; une durée plus petite n'était pas possible étant donnée la périodicité des données disponibles; une durée plus grande risquait de masquer certaines variations importantes. La durée n_1 de la période considérée avant l'instant de la prévision t est fixé à 19 demi-heures. Ce choix est le résultat de la confrontation entre un certain nombre de tests numériques et l'avis des experts en gestion du trafic autoroutier. Ainsi, en fixant h l'horizon de la prévision à 1 heure (i.e. 2 fois θ), le vecteur profil relatif au jour j à l'instant t, $P[j,t]$, a 22 composantes :

$$P[j,t] = [P^t_{ji} \,|\, i = t-19, t+2]$$

$$\text{où} \quad P^t_{ji} = q^t_{ji} / [\textstyle\sum_m q^t_{jm} \,|\, m = t-19, t+2]$$

Ces profils, rappelons-le, sont calculés uniquement avec les débits du point de prévision.

Avant de procéder à la classification nous avons procédé à un certain nombre d'analyses des données historiques et nous avons déduit de la forme et de la structure des déplacements en Juillet et en Août à un regroupement des données selon les mois; ceci est dû en partie au facteur "socio-économique" du voyage en ces mois de congés en France. Comme nous l'avons déjà signalé (cf.II.1.3) le regroupement saisonnier des données de trafic n'aboutit pas, toujours, à une distinction strictement mensuelle des données. Dans une étude effectuée (cf. [10]) sur les données de circulation du réseau national on arrive à cette conclusion que les 12 mois du trafic en rase campagne en France se scindent uniquement en trois ou quatre groupes de mois.

Pour la classification des profils, après plusieurs tests de sensibilité nous avons conclu que pour les données du type routier ou autoroutier en rase campagne, la classification faite sur les profils demi-horaires n'apporte pas plus d'information que la classification sur les profils agrégés sur une heure. Ainsi les objets à classer sont les vecteurs profils ayant chacun

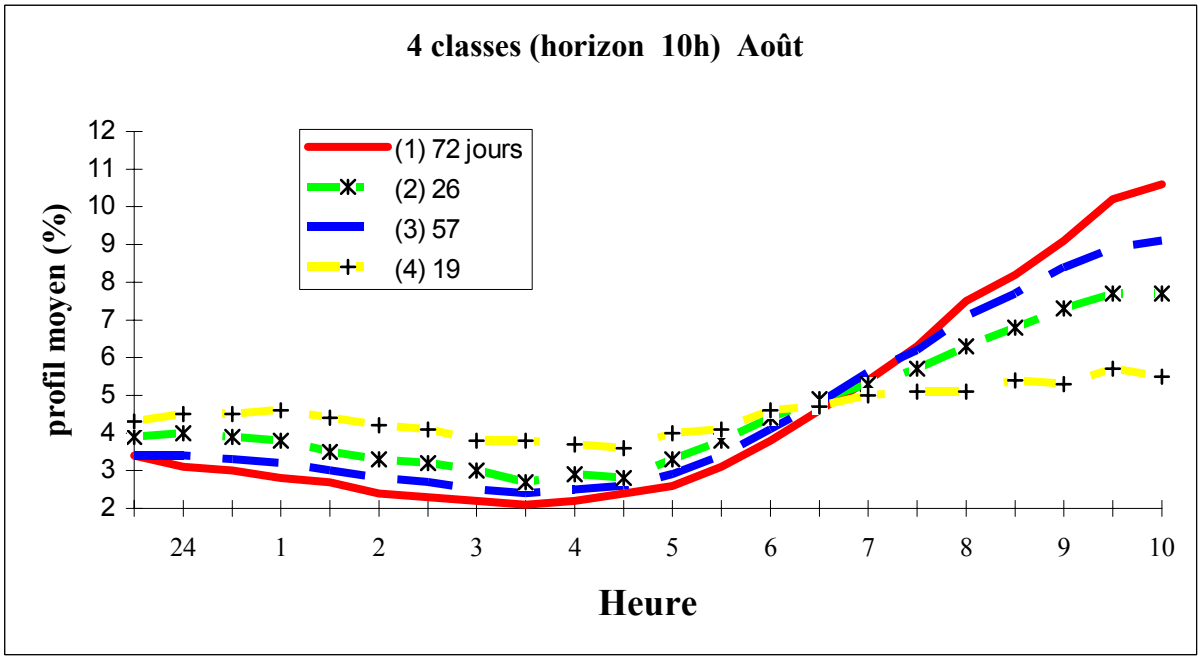
11 composantes; une classification est effectuée à chaque heure (i.e. t=heure) et pour chacun des deux mois (Juillet et Août). Par contre, pour que à l'étape de prévision le critère de choix de classe soit bien précis, les profils moyens des classe P[c,t] sont à nouveaux exprimés par des vecteurs de 22 composantes:

$$P[c,t] = [\, p^t_{ci} \,|\, i = t-19, t+2]$$

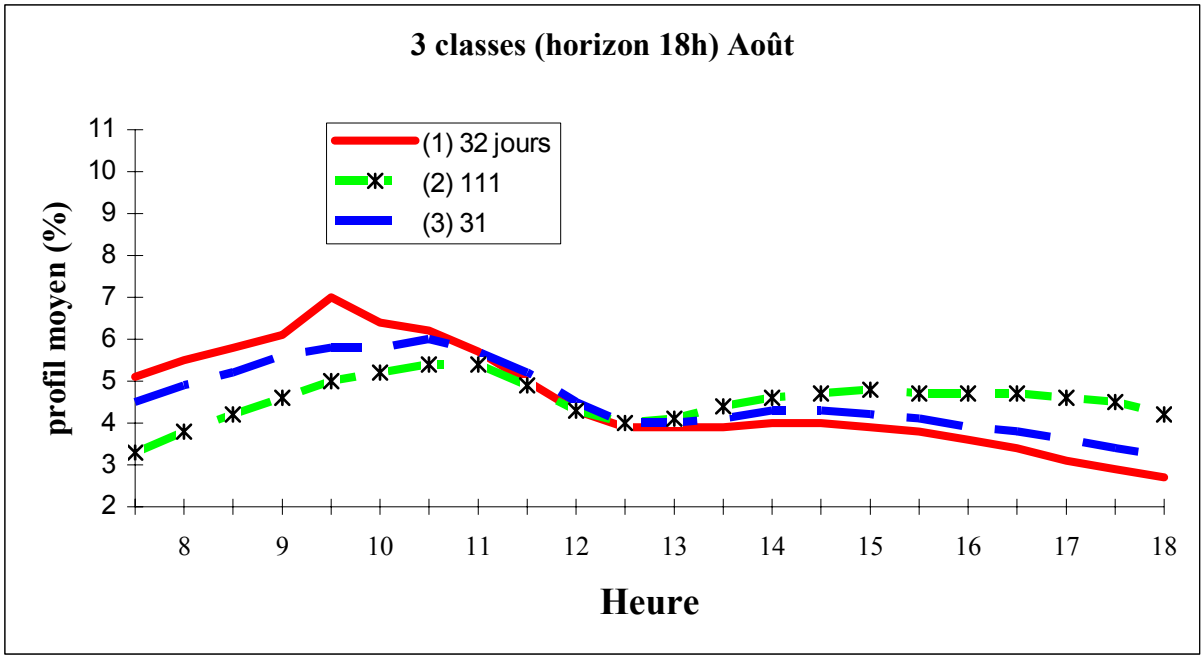
Où $p^t_{ci} = [\sum_j p^t_{ji} \,|\, j \in c] / e_c$

e_c étant l'effectif de la classe c

Dans les deux graphiques qui suivent on donne deux exemples de classifications dont le premier est destiné à la prévision à 10h (i.e. t+2 = 20) pour le mois d'Août et le deuxième est pour le même mois et à 1'horizon de 18h (i.e. t +2 = 36).



La classification de 10h apporte 72% de l'information (variance) totale (4942), les profils moyens des classes ont un comportement parallèle jusqu'à 5h et au delà se distinguent par leur variation dans le temps. Notamment les évolutions très différentes entre 9h et 10h impliquent que les prévisions faites à 9h pour 10h dépendent de l'appartenance à la classe.



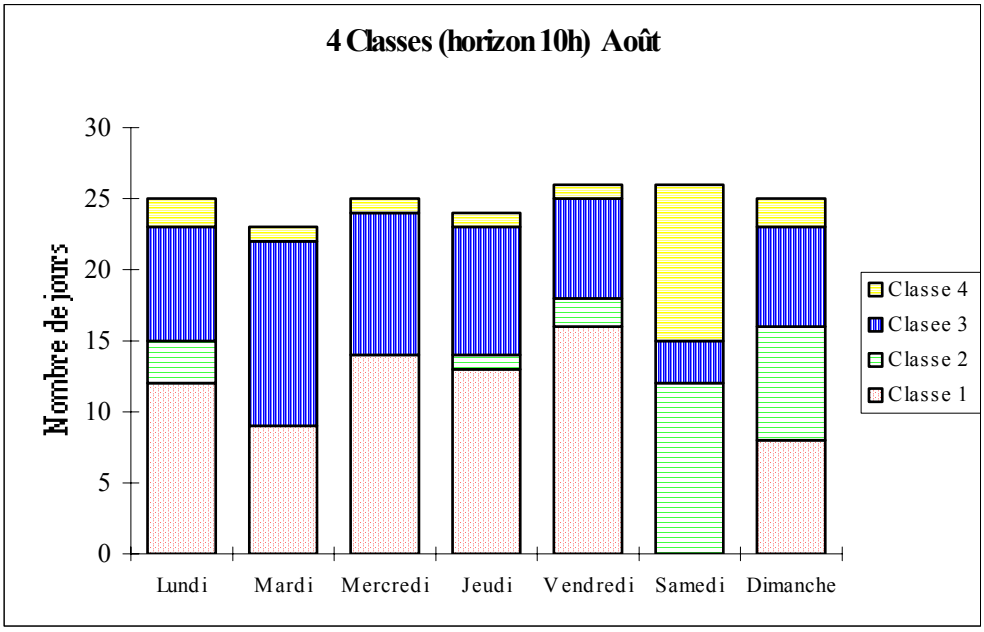
La classification de 18h produit uniquement trois classes. La raison en est que la variance totale des profils (3279) est beaucoup plus faible que celle de 10h, ce qui signifie que dans le cas de 18h les 174 profils sont beaucoup plus homogènes. Par ailleurs, la part de l'information contenue dans cette classification est de l'ordre de 55%. Les courbes des profils moyens se croisent entre 12 et 13h; mais elles divergent avant et après.

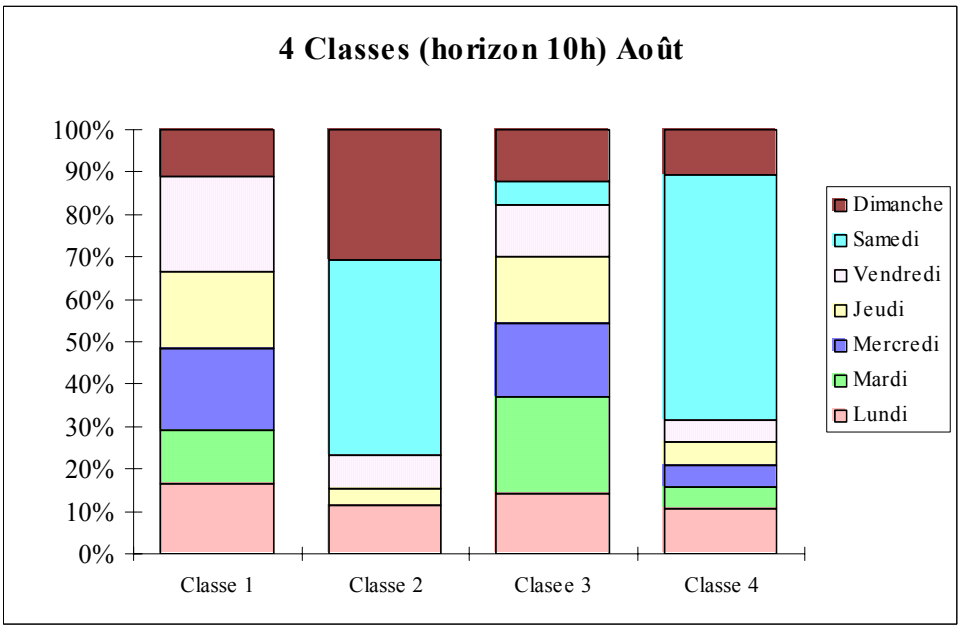
Le nombre de classes par heure varie entre 3 et 5, au total pour les deux mois de calibrage nous avons obtenu 192 classes.

III.3 Le croisement entre la classification mathématique et la classification calendaire

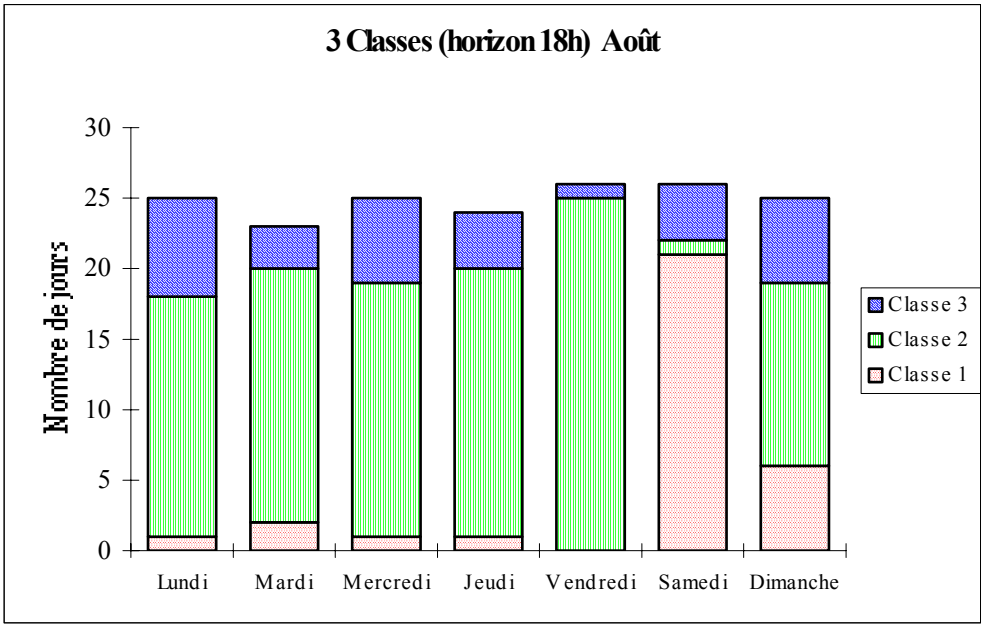
Pour mieux apprécier l'apport de la classification des profils par rapport à la classification calendaire pour les 2 exemples ci-dessus, on a croisé leurs classes avec les 7 jours de la semaine :

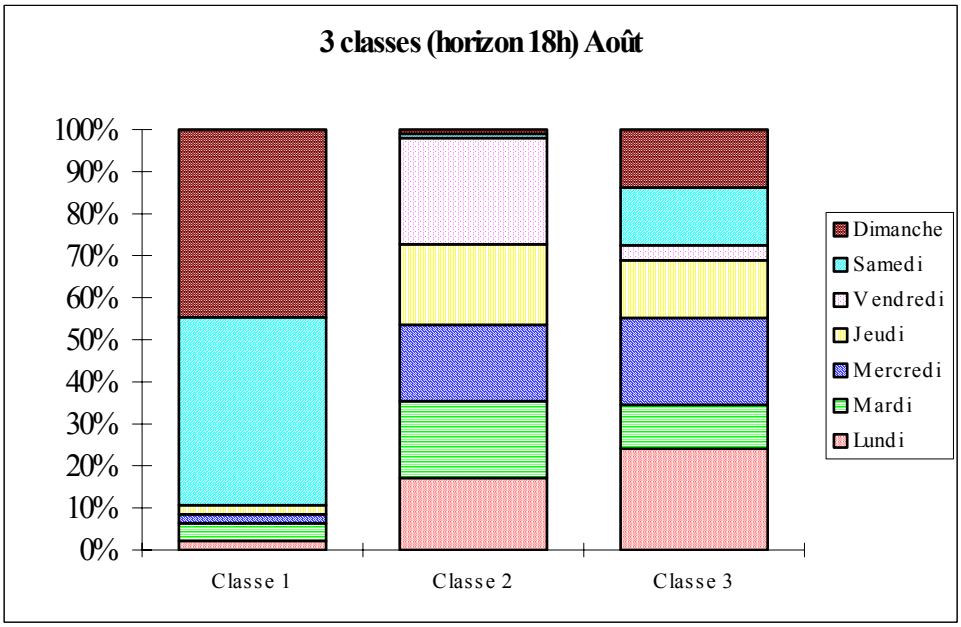
Dans le premier exemple (4 classes pour l'horizon l0h), la classe n°2 est composée essentiellement des Samedis et des Dimanches. Les Vendredis sont bien répartis entre les classes sauf dans la n°2 où ils ne figurent pas. La classe n°4 est composée un peu plus de





moitié des samedis et le reste les autres jours. Les classes n°1 et n°3 se sont à 90% remplies des jours ouvrables; dans la n°1 on trouve quelques dimanches mais aucun Samedi; dans la n° 3 il y a un peu des deux.





Dans le deuxième exemple (3 classes pour l'horizon 18h), la classe n°1 contient presque tous les week-ends, Dans la classe n°3 on trouve tous les jours, la n°2 est la même sauf qu'on n'y trouve aucun vendredi. On constate ainsi que la classification automatique va au delà d'une simple classification calendaire; en particulier elle varie en fonction de l'heure de la journée.

Par exemple les jours de week-end ne se ressemblent pas tous et selon le profil du trafic ils se regroupent différemment.

III.4 Le calibrage des modèles

Pour chaque classe ainsi constituée nous avons calibré un modèle de régression linéaire; le nombre de coefficients varie entre 6 et 1, selon l'effectif des classes, Pour la classe ayant au moins 50 éléments nous avons choisi un modèle avec 6 coefficients et pour les classes constituées d'au plus de 10 éléments le nombre de coefficients a été limité à un.

Dans cette application nous avons testé 3 types de modèle :

Modèle 1	les variables explicatives proviennent, uniquement, de Beaune; les coefficients peuvent être négatifs ou positifs
Modèle 2	les variables explicatives proviennent de Beaune et d'Avallon; les coefficients sont contraints à être positifs
Modèle 3	les variables explicatives proviennent de Beaune, d'Avallon, de Baume-les-Dames et de Langres; les coefficients sont contraints à être positifs

D'après les tests effectués dans le cadre de cette application ce dernier type de modèle produit les meilleurs résultats; ceux-ci sont détaillés plus loin. Pour l'horizon de prévision h+2 (exprimé en θ) et selon le nombre de coefficients, les variables explicatives proviennent du point de prévision (i.e. Beaune) et d'un, de deux ou des trois amonts (cf. schéma suivant). Le temps de parcours moyen entre chaque amont et Beaune est estimé à 60 minutes (σ = 20). La contrainte de positivité se réalise par une méthode de pas à pas (cf. annexe B) et conduit à annuler certains coefficients, c'est pourquoi le tableau suivant ne définit que l'ordre des coefficients.

Variables explicatives du modèle n°3 (horizon h+2 exprimé en demi-heure)

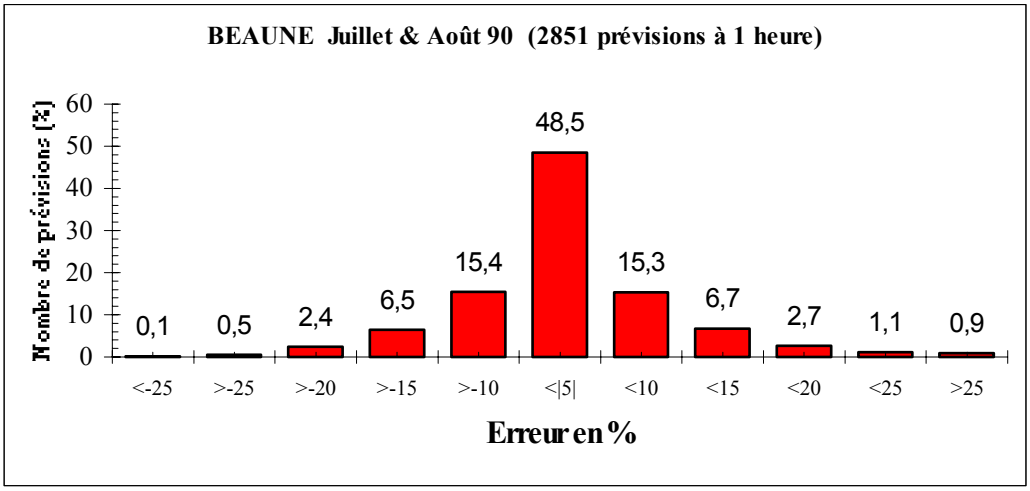
Nb. Coeffs.	Beaune h	Avallon h	Baume-ld h	Langres h	Beaune h-1	Langres h-1
1	*					
2	*	*				
3	*	*	*			
4	*	*	*	*		
5	*	*	*	*	*	
6	*	*	*	*	*	*

III.5 Le résultat de l'application d'ATHENA (modèle n°3) en Juillet et Août 90

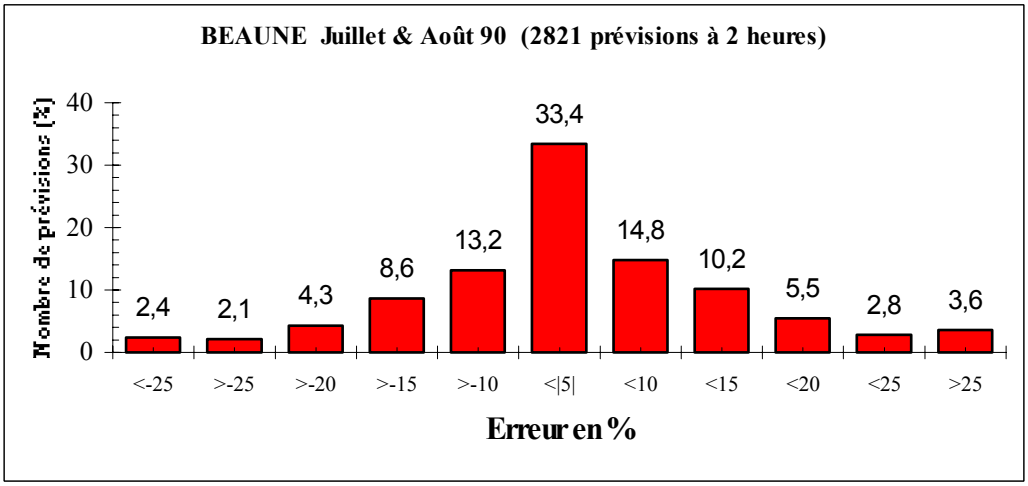
Le modèle a été appliqué pour 2851 demi-heures dont 1412 en Juillet et 1439 en Août; dans l'ensemble 2.357.429 véhicules ont été détectés par le capteur situé à Beaune (débit moyen de 1.654 véhicule heure), Le modèle ATHENA à l'horizon de 60 minutes a prévu 2.360.751 véhicules (très légère surestimation) et 2.347.379 véhicules à l'horizon de 120 minutes (très légère sous-estimation).

L'erreur globale (quadratique) E_g de prévision à 1 heure s'est élevée à 7,7% (RMSE = 156 véhicules) pour le mois de Juillet alors qu'une simple reconduction du niveau de trafic d'une heure sur l'autre (modèle "plat") conduit à une erreur globale de 14,1%; Pour le mois d'Août l'erreur globale s'est élevée à 10,0% (RMSE=161) contre 15,9% pour le modèle "plat". L'erreur globale de prévision deux heures à l'avance s'est élevée à 10,7% (RMSE=216) en Juillet et 15,8% (RMSE=255) en Août.

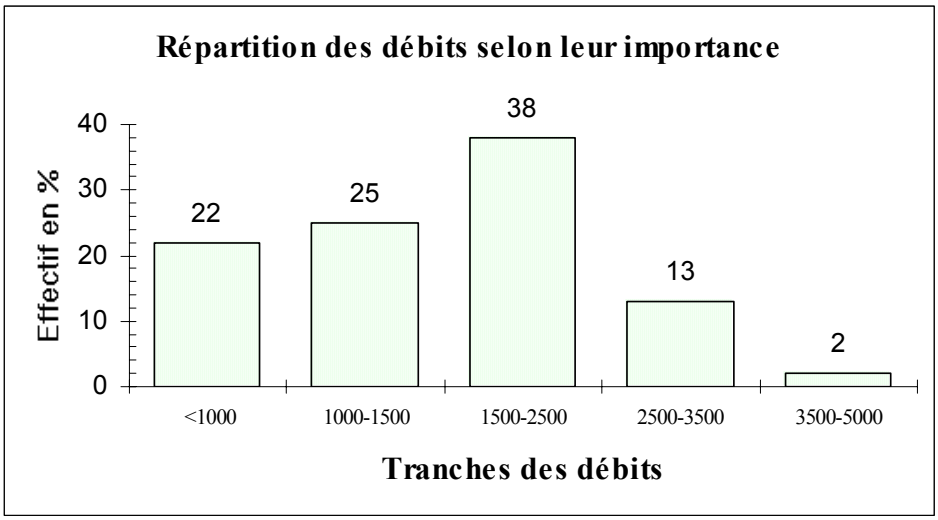
La répartition des erreurs partielles E_p^t de prévision à 1 heure, fait apparaître que dans 79% des cas la valeur absolue de cette erreur est inférieure à 10%. Le schéma ci-dessous montre bien que les erreurs positives et négatives sont réparties d'une manière équilibrée, ce qui signifie qu'il n'y a pas beaucoup plus de sous-estimations que de surestimations.



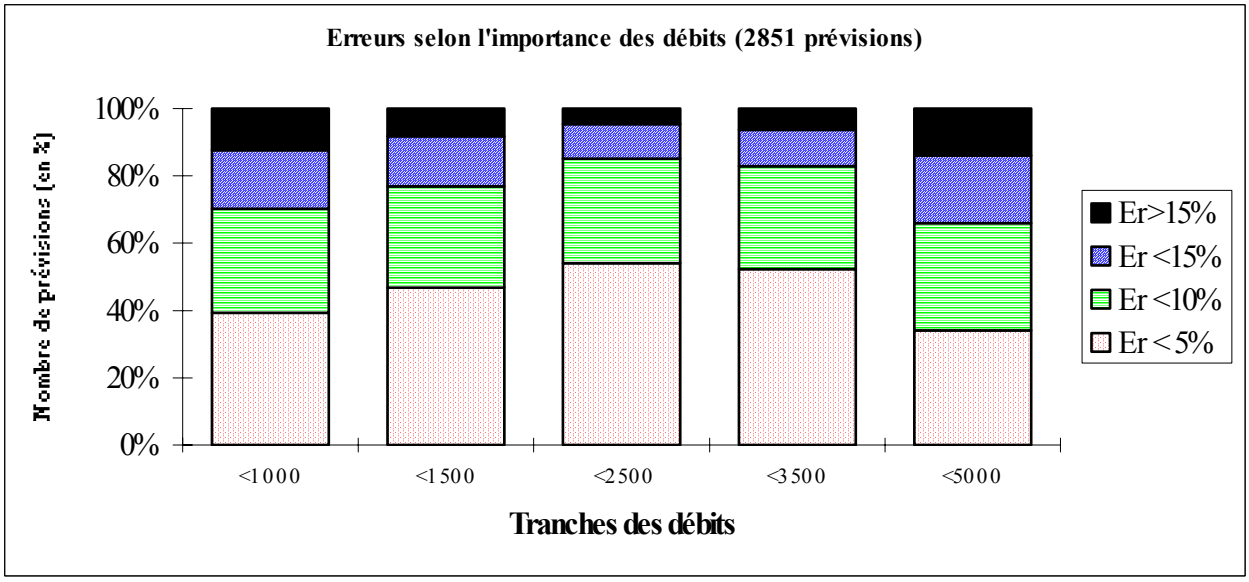
Dans le cas de prévision à 2 heures 61% des erreurs E_p^t sont en dessous de 10%. Le schéma ci-dessous montre un léger déséquilibre entre les sous-estimations et les surestimations, les E_p^t >lO% couvrent 22% des prévisions et les E_p^t <-10% se produisent dans 17% des cas.



Les débits n'ayant pas la même importance, il est intéressant de voir de quelle manière les erreurs E_p^t se répartissent selon les différentes tranches des débits. L'histogramme suivant représente la répartition des débits selon leur importance :



La répartition des erreurs partielles E_p^t (horizon 1 heure) selon ces tranches des débits figure dans le schéma qui suit:



On peut constater que les meilleurs prédictions sont réalisées dans le cas des débits variant entre 1000 et 3500 véhicules/heure ce qui couvre 76% des cas. Les erreurs fortes (>15%) dans le cas de débits faibles (<1000) ne dévalorisent pas le modèle; par contre les erreurs fortes pour les débits forts (>3500) concernent uniquement 2% des prévisions; il s'agit de fluctuations exceptionnelles dans les débits dues aux accidents ou incidents. Dans ce cas, les débits observés en amont ne passent pas forcément en temps prévu au lieu de prévision.

III.6 Explication des erreurs fortes (>15%)

Une partie des erreurs fortes a pu être expliquée par la SAPRR, incidents ou

problèmes de capteur (cf. [6]). En particulier les événements perturbant ont été fréquents au mois d'Août : c'est la raison pour laquelle la performance du modèle est moins bonne en Août qu'en Juillet.

D'après la synthèse de la SAPRR cinq erreurs fortes du mois de Juillet s'expliquent par des accidents :

le 3 Juillet à 13h30

le 8 Juillet à 11h et à 13h

le 20 Juillet à 9h

28 Juillet à 10h (même si le bouchon se termine à 9h20, la prévision de 9h pour 10h est sensible à l'accident).

Le 6 Juillet à 23h le débit d'Avallon est passé de 748 à 1372, le 26 de ce mois et pour le même point à 16h le débit passe de 1254 à 12. Ainsi pour le mois de Juillet, restent inexpliqués 15 cas d'erreur forte (1% de l'ensemble des demi-heures de Juillet),

Pour le mois d'Août, cinq cas sont expliqués par des perturbations du 4 Août entre 5h et 14h30; d'autre part le 5 Août le capteur de Beaune chute de 3268 véhicules heures à 11h à 2840 à 11h30 et 1806 à 12h (le modèle ne peut suivre cette évolution trop rapide); il reste donc, 11 cas d'erreur forte inexpliqués.

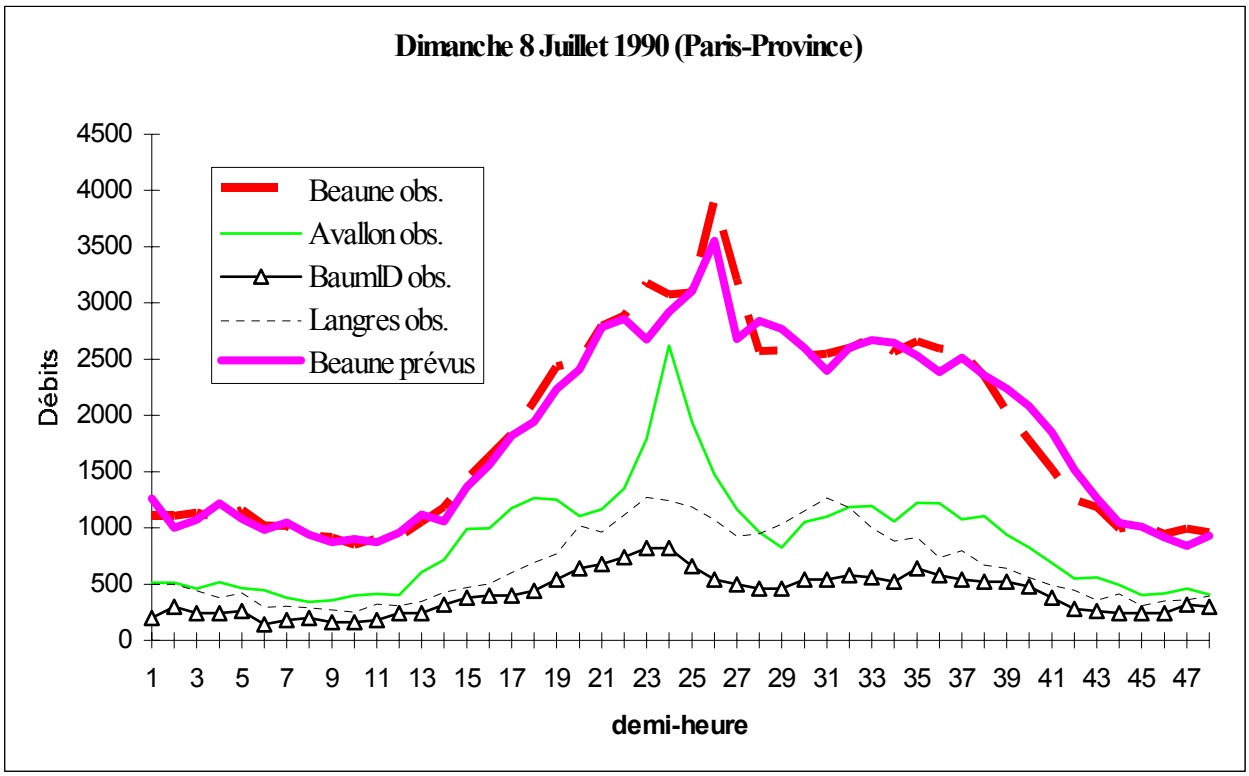
Dans l'ensemble des prévisions effectuées (2851), on trouve uniquement 52 cas non expliqués d'erreur forte, c'est-à-dire 2% ce qui est relativement négligeable.

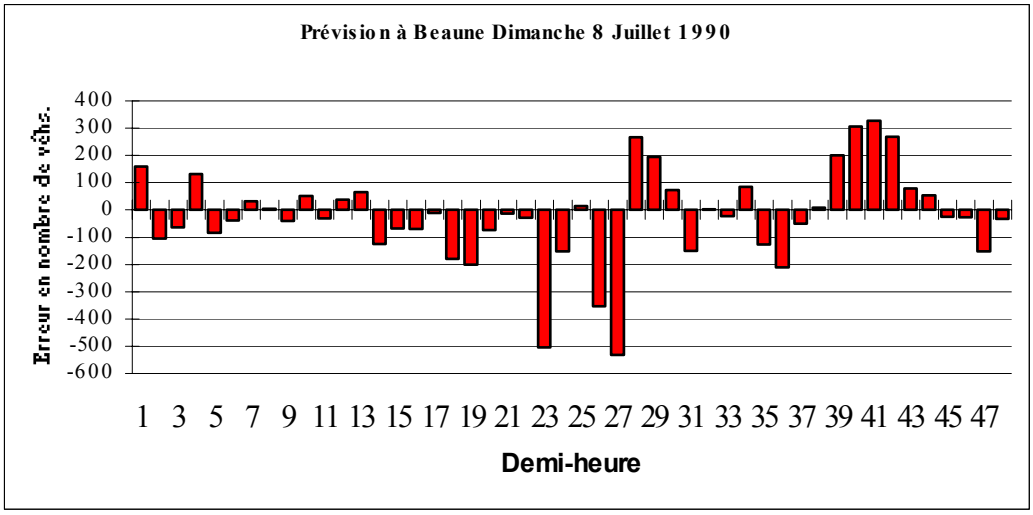
III.7 Erreurs dues aux fluctuations importantes de débits

Le Modèle ATHENA a été appliqué sur 68 journées entières (de 0h à 23h30). L'étude des couples des débits réels et prévus montre qu'en général ces deux se suivent de très près. Le modèle a un comportement quasi-exact au voisinage du maximum de débit journalier, mais il réagit avec du retard aux fluctuations importantes des débits, ce qui explique une partie des erreurs fortes. La raison essentielle de cette faiblesse du modèle est que dans les données historiques ces fluctuations se compensent, ainsi nous n'avons aucune classe de profil qui pourrait englober ce genre de situations.

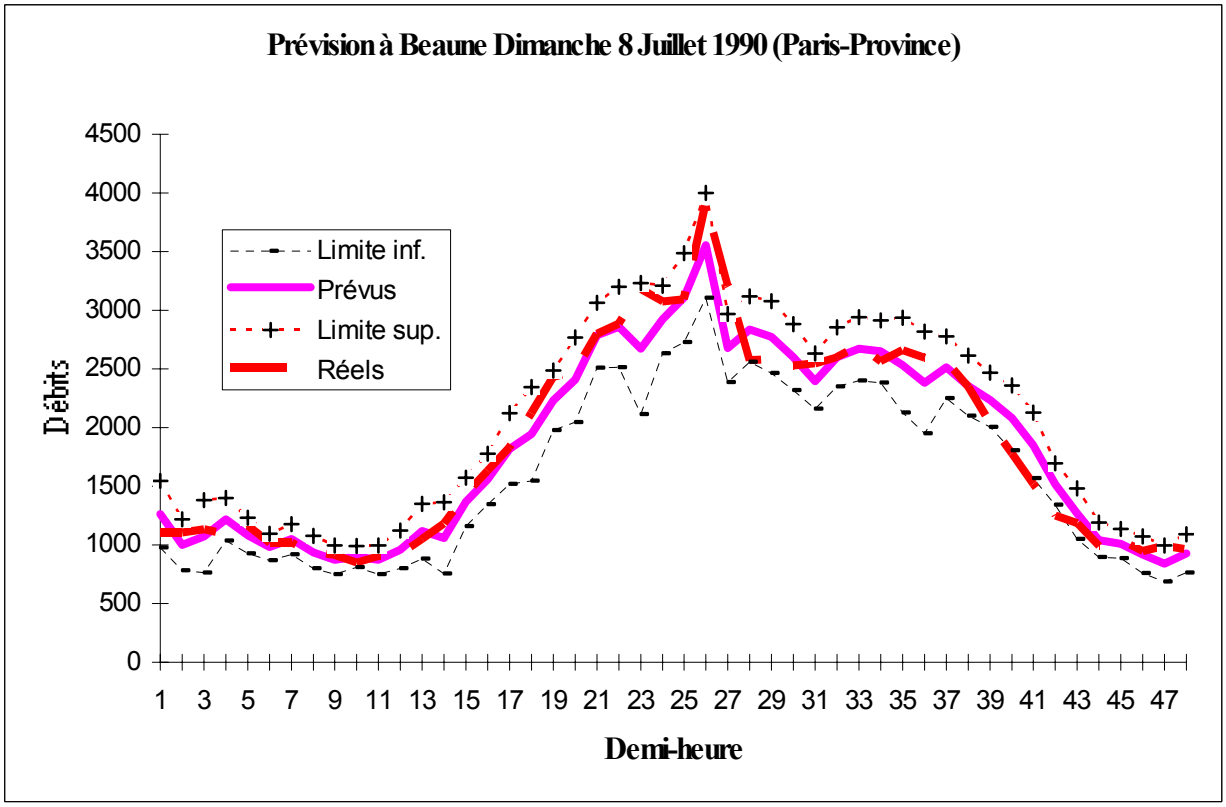
III.8 Un exemple d'une journée de prévision

Pour illustrer cet exemple nous avons choisi un jour qui soit perturbé par un accident; le Dimanche 8 Juillet 1990 un accident avait été signalé par SAPRR à Avallon entre 9h30 et 10h (19 et 20ème demis-heure). D'après les courbes ci-dessous on voit bien que cet accident a répercuté sur la courbe de Beaune de 11 h à 13h (22ème à 26ème demi-heure). Les erreurs les plus fortes sont relatives à cette tranche du temps ; 500 véhicules prévus en moins à la 23-ème demi-heure et 550 en moins à la 27-ème demi-heure (-16%). Pour les autres moments de la journée la courbe prévue suit avec une erreur inférieure à 10% la courbe du trafic réel, sauf entre 19h et 20h où l'erreur est de l'ordre de 17% (300 véhicules). Cette erreur forte s'explique par une baisse exceptionnelle des débits réels; en effet les profils moyens des classes (dans ce cas il y en a trois) ont, dans cette tranche du temps (19h à 20h), plutôt une allure plate que descendante.





Pour chaque valeur prévue nous avons calculé son écart-type (cf annexe B), ce qui permet d'évaluer son intervalle de confiance. Dans le schéma ci-dessous on peut voir la variation de ces intervalles selon les valeurs prédites. On constate ainsi que les deux courbes des intervalles suivent bien la courbe des valeurs prédites. Par ailleurs à quelques exceptions près, la courbe des débits réels est bien à l'intérieur de l'intervalle de confiance. Les deux tranches du temps où la courbe réelle touche les deux limites correspondent exactement aux erreurs très fortes évoquées ci-dessus (i.e. 11h-13h et 19h-20h).



IV CONCLUSIONS

IV.1 L'originalité du modèle

ATHENA est un modèle horaire de prévision du trafic sur une section de route; son originalité réside dans les éléments suivants :

- chaque courbe journalière de débits est scindée en plusieurs segments d'une longueur fixe, chaque segment se terminant par l'horizon de la prévision,
- ces courbes sont ensuite transformées en profil (affranchies de l'effet de taille),
- une typologie est recherchée pour chaque horizon de prévision,
- un modèle linéaire est constitué pour chaque type (classe) de profils,
- dans ces modèles on tient compte aussi bien des évolutions temporelles que spatiales.

Dans cette démarche qui est enrichie des approches développées auparavant dans ce domaine (cf. introduction) nous faisons l'usage d'un processus non linéaire tout en pratiquant des modèles linéaires.

IV.2 Les points forts

Le bilan global de l'application exposée plus haut (cf III), après explication d'un certain nombre d'erreurs fortes, s'élève à 7,3% (critère quadratique) pour le mois de Juillet et 10% pour le mois d'Août. On peut rapprocher ces chiffres avec les résultats de l'application de ce modèle à Vienne sens Province-Paris par ASF pendant l'été 90 (10% d'erreur globale pour 1024 demi-heures de prévision) et de l'application d'ATHENA par la SAPRR durant l'été 1989. Aussi sommes nous en mesure de confirmer que l'objectif initial de ce modèle (rester en deçà d'un seuil de 10% d'erreur globale) est atteint.

IV.3 Les points faibles

L'un des points faibles de ce modèle réside dans le calcul très strict du choix de la classe qui pourrait amener à choisir une classe tandis qu'une autre eût été meilleure. Par ailleurs les exemples d'application montrent une certaine faiblesse du modèle à prédire les très grandes irrégularités dues en général aux congestions. Pour combler ce genre de lacunes nous étudions, actuellement, d'autres approches qui font appel à des modélisations non linéaires utilisées dans l'étude des systèmes dynamiques.

ANNEXES

ANNEXE A

Rappel de quelques notions de la classification automatique

Algorithme des centres mobiles

A.1 L'espace des objets à classer

Prenons E l'ensemble des objets (individus) à classer, n le nombre des objets et m le nombre des variables v définissant un objet. Chaque objet est assimilé à un point dans l'espace à m dimensions; la ressemblance entre les deux points i et i' est définie par la distance euclidienne :

$$d^2(i,i') = [\sum_j (v_{ij} - v_{i'j})^2 \,|\, j = 1,m]$$

où v_{ij} est la j-ème coordonnée de l'objet i

A.2 Le centre de gravité et l'inertie

Le centre de gravité g_k d'une classe k est le point dont chaque coordonnée (il y en a m) est la moyenne, sur les éléments de la classe, de la variable correspondante. L'inertie I_k de cette classe est la moyenne des carrés des distances des objets de la classe k au centre de gravité g_k. On définit l'inertie intraclasse I_W, comme une mesure d'homogénéité des classes :

$I_W = [\sum_k e_k I_k \,|\, k = 1,K]$ où e_k est l'effectif de la classe k et K est le nombre de classes.

L'inertie interclasse I_B mesure le niveau de ressemblance entre les classes;

$$I_B = [\sum_k e_k d^2(g_k, g) \,|\, k = 1,K]$$

où g est le centre de gravité de l'ensemble des objets.

A.3 Le critère de classification

L'inertie totale I de l'ensemble des n points autour du centre de gravité global g est après le théorème de KÖNIG-HUYGHENS, la somme de deux termes :

$$I = I_B + I_W$$

Un critère usuel de classification consiste à chercher la partition, en K classes, telle que I_W soit minimal pour avoir en moyenne des classes bien homogènes, ce qui revient à chercher le maximum de I_B.

A.4 L'algorithme des centres mobiles (cf. [l] & [16])

Il consiste à partir de K points parmi les n points de l'ensemble E (en général tirés au hasard); ces K points définissent une partition de E en K classes. On remplace alors les K points initiaux par les K centres de gravité de ces classes et on recommence. L'algorithme converge rapidement vers la partition optimale, c'est-à-dire celle qui maximise l'inertie interclasse. Cet algorithme est dû à FORGY, mais il existe d'autres variantes comme les méthodes de JANCEY, de MAC QUEEN (k-means) et de BALL and HALL (Isodata). Dans le modèle, ATHENA, nous nous sommes inspirés de la méthode Isodata, de la bibliothèque des programmes MODULAD.

ANNEXE B

Rappel de quelques notions de la régression multiple

B.1 Le modèle

La régression multiple consiste à trouver une relation linéaire entre une variable à expliquer et un certain nombre de variables explicatives:

$$y = X\beta + \varepsilon \quad (1)$$

n est le nombre d'observations, y est le vecteur à n composantes de la variable à expliquer, X est la matrice de n lignes et m colonnes des m variables explicatives observées n fois;

β est le vecteur à m composantes des coefficients et ε est le vecteur à n composantes des résidus.

B.2 L'hypothèse

On fait l'hypothèse que les résidus ne sont pas corrélés entre eux (i.e. leur matrice de covariance est diagonale) et qu'ils ont la même variance σ^2. Les coefficients sont estimés par la méthode de moindres carrés :

$b = (X'X)^{-1}X'y$ où X' est la transposée de X

La variance commune des résidus σ^2 est remplacée par un estimateur sans biais S^2 :

$$S^2 = E[\|y - X\beta\|^2]$$

Remarque: Dans le cas où les variables explicatives sont centrées (i.e. moyenne nulle), la matrice ($X'X$) n'est autre que la matrice de covariance de ces variables.

B.3 L'écart type d'un coefficient

L'écart type d'un coefficient β_j se calcule par :

$\sigma^2_{\beta j} = S^2 (X'X)^{-1}_{jj}$ où $(X'X)^{-1}_{jj}$ est le terme (j,j) de la matrice $(X'X)^{-1}$

Sous l'hypothèse de normalité des coefficients, $\beta_j \sim N(0, \sigma_{\beta j})$ l'intervalle de confiance pour un seuil de $100(1-\alpha)\%$ d'un coefficient β_j est :

$\beta_j \pm \sigma_{\beta j} t_{(1-\alpha/2)}$ avec n-m-l degrés de liberté pour le test Student.

B.4 L'écart type d'une valeur prédite

A l'étape de la prévision, à l'aide de relation (l), pour un nouveau vecteur X_0 des variables explicatives, on prédit la valeur y_0. La variance de cette dernière s'obtient:

$\sigma^2_{y_0} = x'_0 (X'X)^{-1} x_0 S^2 + S^2$

L'intervalle de confiance de la valeur prédite y peut s'obtenir par

$y \pm \sigma_{y_0} t(1-\alpha/2)$ avec n-m-l degré de liberté pour le test Student.

B.5 Le test du coefficient de détermination

Le coefficient de détermination R2 (R est le coefficient de corrélation multiple) est le rapport de la variance expliquée par le modèle sur la variance totale de la variable y; sa valeur peut être soumise à un test de Ficher :

$(n-m-1)R^2 / m(1-R^2) \sim F(m, n-m-1)$

Ceci permet de mieux juger la qualité de l'ajustement.

B.6 La contrainte de positivité des coefficients

Dans la méthode descendante de régression pas à pas, dans le premier pas on introduit toutes les m variables explicatives. Dans le deuxième on retire la variable explicative la moins significative et on refait la régression avec m-l variables; le dernier pas du modèle de régression a une seule variable. En général la variable qui est éliminée à chaque pas, est celle qui provoque la diminution la plus faible des R^2 ,dans ATHENA c'est celle qui a le plus petit coefficient négatif et l'itération s'arrête quand tous les coefficients sont positifs.

ATHENA MODE D'EMPLOI

ANNEXE C

MODE D'EMPLOI D 'ATHENA

Cl.	LES PARTICULARITES DE CETTE VERSION	1
C2.	PARTIE TEMPS REEL (PREVISION)	4
	C2-1. DEROULEMENT DES OPERATIONS	4
	C2-2. PROCEDURES INDEPENDANTES	10
	C2.3. LES FICHIERS A PREPARER PAR L'UTILISATEUR	12
C3.	PARTIE TEMPS DIFFERE (CALIBRAGE)	18
	C3-1. DEROULEMENT DES OPERATIONS	18
	C3.2. LES FICHIERS A PREPARER PAR L'UTILISATEUR	22

C1. LES PARTICULARITES DE CETTE VERSION

1. Cette version peut faire à chaque demi-heure quatre prévisions pour les quatre prochaines demi-heures.

2. La période de prévision est délimitée par celle des données historiques.

3. Cette version propose les modèles suivants :
   - B1 le trafic de la section est expliqué uniquement par le débit de cette section, les coefficients sont positifs ou négatifs ;
   - B2 le trafic de la section est expliqué uniquement par le débit de cette section, les coefficients sont positifs;
   - B3 le trafic de la section est expliqué par les débits de cette section et une autre située en amont, les coefficients sont positifs ;
   - B4 le trafic de la section est expliqué par les débits de cette section et deux/trois autres situées en amont, les coefficients sont positifs ;

4. Pour que la prévision soit possible à un moment donné, il faut que le modèle ait en mémoire au moins les débits horaires des 22 demi-heures précédentes du point de prévision et ses amonts.

5. La saisie des données est automatique pour certains cas.

6. A chaque moment on peut consulter le tableau des 24 dernières prévisions et dresser les courbes des débits réels et prévus.

7. L'ensemble des prévisions est enregistré dans un fichier intitulé "ARCHIV".

8. Un programme "stat" indépendant est fourni pour dresser un tableau ré- capitulatif à partir du fichier "ARCHIV".
9. On peut utiliser le fichier "ARCHIV" pour refaire des prévisions.
10. Ce modèle est destiné à une utilisation continue, pour une utilisation discontinu, il faut rentrer les débits des 48 derniers demi-heures

C2. PARTIE TEMPS REEL (Prévision)

C2-1. Déroulement des opérations

A00) Taper catch /Pn /Oi" et Entrée (ceci identifie l'imprimante)

/Pn = no d'identificateur (voir page suivante)

/Oi = pour imprimer à l'italienne (mettre une grande feuille)

Cette commande est faite une fois après le redémarrage de l'ordinateur.

A0l) Taper " Athena "

affichage de l'Ecran titre

ATHENA propose le premier modèle, si l'utilisateur est d'accord, il fait (Entrée) sinon il choisit un autre modèle

affichage du MENU avec les options suivantes :

A - Saisie automatique ou par ARCHIVE

X - Saisie automatique (exemple :SNRD - M II - Xtedi) S - Saisie

G - Graphique

C - Capacité à modifier

L - Liste et modifications

P - Prévision

T - Tableau récapitulatif M - Modèles

Q - Quitter avec sauvegarde B - Quitter sans sauvegarde

IDENTIFICATION DES IMPRIMANTES
POUR LA PROCEDURE CATCH

1. *AT&T 477*
2. *Canon BJ-30*
3. *Citizen MSP-10/15, 20/25, 40/45, 50/55*
4. *Citizen HQP-45, MSP-35*
5. *Epson Série EX 293*
6. *Epson Série FX*
7. *Epson Série LQ*
8. *Epson Série LX*
9. *Epson Série MX*
10. *Epson Série RX*
11. *HP LaserJet*
12. *HP LaserJet Plus* ou Série II
13. *HP Thinkjet* (en mode HP)
14. *HP Thinkjet* (en mode Alterné)
15. *IBM Graphics* ou *Proprinter*
16. *IBM Colorjet*
17. *NEC P5, P6, P7*
18. *Okidata 84 step 2, 92, 93, 192, 193, 292,*
19. *Okidata 2410*
20. *Okidata* en mode IBM
21. *Olivetti PR-15/17/38*
22. *Panasonic KX-P1080i, P1091i, P1093*
23. *Panasonic KX-P1524, P1592, P1595*
24. *Star Micronics NB 24-10, 24-15, 15*
25. *Toshiba P1340, P1350, P1351*
26. *Toshiba P321, P341, P351*

SAISIE (manuelle)

S02) Option S : taper s
On verra s'afficher "entrez l'horodate actuel (JJ/MM/HH/MN) "

S03) Taper l'horodate dans l'ordre indiqué (séparé par un /)

S04) A la question de vérification taper NON, si pas d'accord (retour SO3) si OK (Entrée)

SO5) L'horodate et les derniers débits mémorisés pour l'ensemble des points (point de prévision et son (ou ses) amont(s) vont être affichés. On doit saisir les débits (Véh/h) par demi-heure et par point depuis l'ancien horodate jusqu'à l'horodate actuel (maximum 48 débits pour chaque point). Après chaque demi-heure l'utilisateur a la possibilité de vérifier les dernières données entrées. A la fin de la saisie on retrouve l'écran MENU.

Remarque 1 : la prévision est impossible si l'utilisateur ne dispose pas pour la période indiquée de débits pour le point de prévision. Par contre dans le cas où les débits d'un ou des points amont sont manquants, l'utilisateur saisira les débits nuls pour ce(s) point(s) et il fera attention à ne pas appliquer pour cette période le(s) modèle(s) mettant en jeu ce(s) amont(s).

Remarque2 : Pour avoir une meilleure prévision, les PK des points doivent correspondre à ceux pris en compte dans le calibrage d'ATHENA.

SAISIE par ARCHIVE

A02) Option S : taper A

On verra s'afficher "entrez l'horodate actuel (JJ/MM/HH/MN)"

A03) Taper l'horodate dans l'ordre indiqué (séparé par un /)

Le modèle cherchera les données nécessaires dans le fichier "ARCHIV" et effectuera les calculs de "prévisions"

GRAPHIQUE

G02) Option G : taper g

On verra s'afficher les deux courbes débits réels et prévus. Pour impri- mer ces courbes, il suffit d'appuyer sur la touche "imprimer écran (Prt Screen)". Si l'imprimante est en noir et blanc, les couleurs des courbes sont imposées par ATHENA. Si l'imprimante est en couleur, on peut choisir les couleurs. Ce choix s'effectue avant de lancer ATHENA à l'aide de la procédure "SELCOUL". L'utilisateur a la possibilité d'enregistrer ces images sur une disquette, pour cela il faut, avant de lancer ATHENA, appeler la procédure "GARIMAGE".

CHANGEMENT DE CAPACITE

C02) Option C : taper c

Cette option du MENU permet de changer la capacité maximum des points de comptages.

LISTE ET MODIFICATION

L02) Option L : taper l

On verra s'afficher pour chaque point la liste des derniers 48 débits mémorisés.

L03) L'utilisateur a la possibilité de modifier un ou plusieurs débits. D'abord il choisit l'heure et la minute (sépare par un /) et ensuite il donnera la nouvelle valeur. A la fin des modifications ou consultations on retrouve l'écran MENU.

PREVISION

P02) Option P : taper p

Si on utilise cette option avant l'option "m" (modification du modèle), ATHENA opte pour le modèle choisi au départ (voir A01) Sinon la prévision sera faite avec le nouveau modèle. Pour chaque demi-heure on voit afficher les informations suivantes :

- Numéro de la classe la plus proche
- Heure et minute
- Ecart-type de prévision pour les 4 demi-heures suivantes

A la fin de la prévision on voit s'afficher un tableau donnant la valeur de débit en véh/h pour les 4 demi-heures ainsi que leur écart-type. Pour avoir la fourchette de prévision il suffit de calculer

limite inférieure = débit prévu - 2 fois écart-type
limite supérieure = débit prévu + 2 fois écart-type

Attention cette fourchette n'est valable que si les débits saisis sont ex- primés en véhicule/heure (Véh/h).

TABLEAU_RECAPITULATIF

T02) Option T : taper t

On voit s'afficher les 48 dernières prévisions et leurs écarts relatifs et absolus par rapport aux débits réels. A la fin on trouve dans un encadré l'écart quadratique moyen évalué sur 48 demi-heures. En vue de faciliter l'impression éventuelle de ce tableau (avec la touche Prt Sc, l'affichage se fait en trois parties. On passe aux parties suivantes en faisant Entrée. Un fichier "ATHENA.RES" identique à ce qui défile à l'écran est créé; il sera écrasé à la session suivante.

MODELES

M02) Option M : taper m

Avec cette option on peut choisir un des trois modèles cités plus haut.

M03) Si on est d'accord avec le modèle proposé (affiché) on tape Entrée si- non, on tape non, et on choisit un autre modèle. Une fois que le modèle est désigné, l'écran MENU sera affiché.

M04) Pour faire la prévision avec ce nouveau modèle, on choisit l'option "p".

ARRET_AVEC_SAUVEGARDE

Q02) Option Q : taper q

Arrêt du programme avec sauvegarde des résultats dans le fichier "ARCHIV".

ARRET_SANS_SAUVEGARDE

B02) Option B : taper b

Arrêt du programme (en cas de fausse manoeuvre sans sauvegarde des résultats).

C2-2. ***Procédures indépendantes***

1. STATISTIQUE PERIODIQUE - Procédure "STAT"

Indépendamment du logiciel ATHENA, l'utilisateur a la possibilité d'établir, pour une période et un modèle donnés, des tableaux journaliers, des courbes "débits réels", "débits prévus" et pour l'ensemble de la période les écarts quadratiques et des histogrammes. Pour faire ceci il suffit de taper "stat" et répondre aux questions. Les tableaux issus de cette procédure sont automatiquement enregistrés dans un fichier nommé "stat.res". L'impression des courbes se fait, de la même manière que dans le cas de l'option G du MENU (voir ci-dessus).

Les tranches de débits choisies par le logiciel peuvent être changées de la manière suivante : l'utilisateur, à l'aide d'un outil "éditeur de texte", changera à son grès les 6 derniers chiffres de l'unique ligne du fichier "TRANCHE". Le premier chiffre (0 ou 1) ne doit pas être changé.

N.B. : Cette procédure tient compte du dernier fichier "ARCHIV" existant.

2. CHOIX DE L'IMPRIMANTE - Procédure : "IMPRIM"

L'utilisateur a la possibilité de choisir entre deux sortes d'imprimantes : "noir et blanc" ou "couleur". Ce choix se fait avant le lancement d'ATHENA et à chaque fois qu'il décide de changer de type d'imprimante. Pour ce 'faire, il lui suffit de taper "imprim" et répondre à une question. Dans le cas de l'imprimante "noir et blanc" les couleurs de l'affichage et de l'impression des graphiques sont imposées par ATHENA. Quand l'imprimante est "couleur", l'utilisateur a la possibilité de fixer les couleurs lui- même à l'aide de la procédure "SELCOUL ".

3. COULEUR DES GRAPHIQUES - Procédure "SELCOUL "

Cette procédure permet à l'utilisateur de choisir de meilleures couleurs d'impression pour son imprimante couleur. Pour ce faire, il tape "sel- coul" et, dans une première étape, il note les identificateurs des couleurs (des nombres compris entre 1 et l5), ensuite il répondra aux questions.

4. RECALCULER LES PREVISIONS - Procédure "PREV0"

Si, pour différentes raisons, l'utilisateur veut recalculer les dernières prévisions, il lui suffit de taper "prev0" avant de lancer ATHENA.

5. REMISE A ZERO DES FICHIERS "DEBITS" - Procédure ” ZERO ”

Cette procédure remet les fichiers "debits courants" en leur état initial. Elle est surtout utile quand on désire faire de la "prévision" pour une période antérieure à la dernière prévision mémorisée.

6. COPIE FICHIER "ARCHIV" - Procédure "COPARCH"

Pour ne pas créer un fichier "ARCHIV" trop grand, nous conseillons de copier de temps en temps (par exemple une fois par mois) ce fichier sur un autre fichier. Pour cela on peut utiliser la procédure

"COPARCH" de la manière suivante :

On tape **coparch** nom du fichier

Cette procédure copie le fichier "ARCHIV" sur le nouveau fichier et ensuite remet à zéro le premier fichier.

7. SAUVEGARDE DES COURBES DE PREVISION

Si l'on veut sauvegarder les courbes dessinées par l'option "graphique" du MENU, il faut procéder de la manière suivante :

- avant de lancer ATHENA on lance la procédure "**GARIMAGE**". Ceci implique que pour chaque courbe affichée par ATHENA on introduise une disquette dans l'unité A,
- pour réafficher les images sauvegardées, on lancera la procédure "**EDIMAGE**".

C2-3 Les fichiers à préparer par l'utilisateur

1) Le fichier décrivant le réseau : "SAISIE"

exemple : (correspondant à Beaune sens Y)

4BEAUNE 1 1 10000 AVALLO 1 1 00000 BAUMEL 1 1 00000 LANGRE 1 1 00000

exemple: (correspondant à Vienne sens W)

2VIENNE l-l 10000 MONTEL l-l 00000

Format:

1)Colonnes 1-3	le nombre de points du réseau (calé à droite)
2)Colonnes 4-9	le nom du capteur en 6 caractères
3)Colonnes 10-11	la valeur 1 (valeur ne pouvant être différente que dans le cas d'une acquisition des données par l'intermédiaire de XTEDI, qui lit les fichiers FIME pour les stations SIRREDO)
4)Colonnes 12-13	1a valeur -1, 1, ou 2 suivant que le capteur est utilisé dans le sens W, dans le sens Y, ou dans les 2 sens
5)Colonnes 14-15	1a valeur 1 si le capteur est un point de prévision, 0 sinon
6)Colonnes 16-19	1e nom du fichier FIME (en 4 caractères) dans le cas de la saisie par XTEDI d'une station SIRREDO, sinon 0000

Les colonnes 4 à 19 sont répétées (sur la même ligne) pour chaque point du réseau (séparé par un blanc).

2) Le fichier décrivant le calibrage effectué pour un point de prévision

nom du fichier : BEAUNE.YM pour le capteur de Beaune dans le sens Y VIENNE.WM pour le capteur de Vienne dans le sens W

Attention: la première partie du nom doit correspondre aux 6 caractères cités dans le fichier **"SAISIE"**

exemple pour Beaune:

```
BEAUNEY 0 24 3 10 6 0 30 55 22 06 09
    1 l+/- 0BEAUNEY
    2  2+++  0BEAUNEY  0AVALLOY
    3  4+++  0BEAUNEY  0AVALLOY  0BAUMELY  0LANGREY
```

exemple pour Vienne

```
VIENNE W O 24 3 10 6 0 30 55 22 0609
        1 l+/- 0VIENNEW
        2  l+++  0VIENNEW
        3  2+++  0VIENNEW  0MONTELW
```

Format

Première Ligne

1) Colonnes l-6 le nom du point (information redondante)
2) Colonne 13 le sens (avec une majuscule)
3) Colonnes 15-16 l'heure de début du calibrage
4) Colonnes 18-19 l'heure de fin du calibrage
5) Colonne 22 le nombre de modèles (2 s'il n'y a pas d'amont, 3 s'il y a au moins 1amont, cf. I), les lignes suivantes décrirons chacune un modèle.
6) Colonne 24-25 le nombre de classes pour l'ensemble des mois étudiés (10)
7) Colonne 28 le nombre de coefficients du modèle de régression (6 pour l'historique demi-horaire et 4 pour horaire)
8) Colonne 30-31 paramètre inutilisé dans ATHENA mais utilise dans le programme STAT associé (le nombre de dixièmes de secondes d'attente entre deux écrans lors de l'édition du tableau donnant les débits réels et les débits observés. La valeur "0" implique l'attente d'un "Retour Chariot")
9) Colonnes 33-34 périodicité des données historiques, en minutes (30 ou 60)
10) Colonnes 36-37 le nombre de caractères d'une ligne du fichier ARCHIV (55)
11) Colonnes 39-40 le nombre de demi-heures d'un profil utilise dans le choix de classe (22)

12) Colonnes 42-43 mois de début du calibrage (éventuellement mois de début-1 si l'on accepte pour ce mois le calibrage fait pour le mois de début)

13) Colonnes 45-46 mois de fin du calibrage (éventuellement mois de fin+1 si l'on accepte pour ce mois le calibrage fait pour le mois de fin)

Ligne décrivant un modèle (une ligne par modèle cf. ci-dessus)

1) Colonne 3 numéro du modèle

2) colonne 6 nombre de capteurs concernés par le modèle égale au nombre de points "amont" +un (pour le point de prévision)

Ensuite on a pour chaque capteur concerné :

3) Colonnes 7-9 un commentaire qui sera édité lors des prévisions en temps rée1 pour permettre le choix du modèle.

4) Colonnes 10-12 valeur non utilisée (0).

5) Colonnes 13-18 sur 6 caractères l'intitulé du lieu.

6) Colonne 19 le sens (en majuscule) Y ou W.

Nota : le programme STAT demande, pour les points amont, l'existence de fichiers analogues comportant au moins la première ligne; par contre les lignes suivantes (décrivant les modèles) peuvent ne pas exister.

3) Les Fichiers **VIENNE.WW . . . . . BEAUNE.YY**

Pour chaque lieu (orienté) on crée un fichier suffixé par YY ou WW. Ces fichiers contiennent la capacité de la section (les prévisions sont écrêtées à cette capacité et les débits courants pour les 49 dernières demi- heures connues.

Format

Première ligne

La capacité de la section sur 8 chiffres (les prévisions sont écrêtées à cette capacité)

Deuxième à la 50ème ligne

1) Colonnes l- 3 Jour
2) Colonnes 4 – 6 mois
3) Colonnes 7- 9 heure
4) Colonnes 10-12 minute (valeurs possibles : 0 ou 30)
5) Colonnes 13-17 débit (le double du débit de la demi-heure se terminant à la date indiquée)

Nota : Lorsque l'on veut faire une saisie de débits, le logiciel, avant la mise à jour de ce fichier, compare la date courante à la date figurant sur la dernière ligne du fichier:

- s'il y a égalité, il n'y a pas de saisie à faire
- si la date courante est supérieure à la date du dernier débit enregistré, le logiciel demande les données intermédiaires (sans remonter au delà de 49 données)

- si la date courante est inférieure à la date du dernier débit enregistré il y a contradiction (le temps se déroule toujours dans le même sens) Cela peut arriver si on s'aperçoit d'erreurs ou si on veut tester à posteriori un deuxième modèle et que l'on veut reconstituer rétrospectivement les prévisions; dans ce cas il faut faire la procédure "ZERO" pour remettre les fichiers à une date très antérieure.

4) Les fichiers **VILLES.YY ou VILLES.WW**

Ces fichiers servent à initialiser les fichiers BEAUNE..YY ou VIENNE.WW, ceci par l'intermédiaire de la procédure "ZERO".

C3- PARTIE TEMPS DIFFERE (Calibrage)

C3.1 Déroulement des opérations

OBJECTIF : Calculer des profils de débits, classer les profils et établir des modèles de régression à partir des données historiques.

OUTIL : Un logiciel d'éditeur de textes (exemple : MOD)

AVERTISSEMENT : Un logiciel de calibrage d'un modèle ne peut jamais être à 100% automatique. C'est pourquoi l'utilisateur doit avoir une connaissance parfaite des données historiques, une notion de la statistique et la maîtrise du modèle ATHENA, pour pouvoir intervenir dans toutes les étapes de calculs.

A- OPERATIONS COMMUNES

Al- Préparation du fichier "débits" du point de prévision

A2- Epuration du fichier débit du point et tri selon le mois, l'année, le jour

A3- Préparation et épuration des fichiers "débits" amont et mise en concordance avec le fichier du point de prévision.

Remarque A : les trois étapes se déroulent séparément.

B- OPERATIONS PAR HEURE (24 fois)

B4- Calcul des profils horaires

B5- Classification des profils

B6- Vérification des résultats (corrections éventuelles)

B7- Régression

B8- Vérification des résultats (corrections éventuelles).

Remarque B : Les opérations B4, B5, B7 sont regroupées en une seule procédure "CALIB90", qui doit être lancée 24 fois.

A- OPERATIONS COMMUNES (Voir l'organigramme pages 25 et 26)

A1- <u>Préparation du fichier "débits" du point</u>

Le contenu d'un fichier "débits" est : l'année, le mois, le jour, le code jour semaine, le débit total du jour et 48 débits demi-horaires (le nombre de véhicules observés durant chaque demi-heure).

Remarque 1 : Lorsqu'un jour n'est que partiellement valable (par ex. jusqu'à l'heure h incluse), on peut le conserver pour les calibrages d'horizon 1..."h-1"de prévision et le supprimer après : en effet le calibrage de l'heure "h" utilise, pour les prévisions à deux heures, les données de l'heure h-l.

Remarque 2 : Lorsque l'on rajoute 1 an, il faut supprimer l'année la plus ancienne pour éliminer l'effet du passé trop ancien et pour ne pas augmenter la taille des fichiers (la procédure DOS de tri échoue si les fichiers sont trop gros).

A2 - Epuration du fichier débits du point et son tri

- Les jours ayant une suite composée de plus de 10 débits égaux sont recherchés par la procédure
- "**SELZER**". Le fichier de sortie est
- "**ZERO.LST**".
- Les jours ayant deux débits successifs tel que leur rapport dépasse le seuil de (300%) seront signalés par la procédure "**SELRAP**". Le fichier de sortie s'appelle "**RAPPORT.LST**".
- Le fichier est trié selon les trois critères (à la fois) : mois, année et jour à l'aide d'un

éditeur de textes ou, par exemple, avec la procédure SORT du système MS-DOS3.3.

- Ajout des débits de la veille en début de mois (procédure AJOUT1).

Chaque mois doit commencer par les débits de la veille. A titre d'exemple quand nous ne traitons que les mois de Juillet et Août, le premier jour du mois de Juillet de chaque année est dédoublé. Pour le mois d'Août de chaque année le dernier jour du mois de Juillet est recopié juste avant la première ligne. Ces jours répétitifs sont distingués par le "00". Cette procédure exige que les mois (Juillet et Août) commencent bien par le premier jour et que les mois de Juillet se terminent bien par le 31.

A3- **Préparation et épuration des fichiers "débits" amonts et mise en concordance avec le fichier au point**.

Dans cette étape pour chaque amont on utilise les procédures
"**SELZER", "SELRAP", et "AJOUT1**".

Ces fichiers sont préparés et triés de la même manière que le fichier du point. **Chaque fichier doit avoir exactement le même nombre de lignes que le fichier point. Les jours manquants auront "-1" comme débit pour les 48 demi-heures**

A3- <u>Intégration d'une nouvelle année à partir du fichier "ARCHIV"</u>

Lorsqu'il s'agit d'intégrer une nouvelle année n + 1 en un point où des prévisions ont été faites, on peut repartir des fichiers ARCHIV et SAISIE d'ATHENA et utiliser la procédure "ANNEUF" celle-ci recopie les débits archivés suivant le bon format, divise les débits par 2 pour qu'ils soient demi-horaires en niveau, propose l'élimination de certains jours (cf. fichier "SEIULS") et ajoute un jour "00" en début de mois ; il met aussi les fichiers enconcordance. Il faut l'utiliser mois par mois ; un fichier est créé par mois et par capteur (ex. Beaune.y08 pour Août). Il faut ensuite intégrer le fichier du mois "n" de l'année "n + 1" dans l'ancien fichier (le nom de Beaune .y) après le 31/m/n et avant le 00/n+1/1984.

B- <u>Opérations par heure (24 fois)</u> (Voir l'organigramme page 27)

Pour lancer ces opérations, il suffit de taper "CALIB90" et de répondre à une seule question concernant l'heure (h) de prévision (de 1 à 24). Si les opérations B4, B5, B7 se déroulent bien on obtient en sortie un fichier qui est désigné par le nom du point, le sens de la circulation et l'heure de la prévision (h-l) (ex : Beaune.l0y). Ce fichier contient (pour les 2 mois) les profils moyens par classe, le nombre d'éléments de chaque classe, les coefficients de régression par modèle et par demi-heure et pour les deux prévisions (à une et à deux heures).

B6- Vérification des résultats des classifications (corrections éventuelles)

Les résultats des classifications sont enregistrés dans le fichier "**SORTIE.OUT**" qui s'agrandit à chaque déroulement de la procédure "**CALIB90**". Pour éviter d'avoir un fichier trop important, on peut renommer (ou éliminer) ce fichier et ensuite faire "SORT0" afin de l'initialiser.
Les vérifications ci-dessous s'effectuent à la fin de la procédure "**CALIB90**" et à l'aide du fichier "**SORTIE.OUT**" :

- vérifier le nombre et l'effectif des classes
- vérifier la composition des classes à effectif faible
- vérifier les jours les plus éloignés de chaque classe
- vérifier l'information (taux d'inertie) apportée par la classification
- si nécessaire corriger le fichier débits et relancer pour la même heure la procédure "CALIB90". Par exemple supprimer les jours (ou une partie des jours) d'une classe ayant moins de 10 éléments.

Remarque :
Ces suppressions sont très probablement pertinentes pour les heures ultérieures proches ; les classifications changeant peu d'une heure sur l'autre. On peut éventuellement réintroduire les jours supprimés pour les calibrages suffisamment postérieurs (après 4 h).

B8 - Vérification des résultats des régressions (corrections éventuelles)
Au cours de l'exécution de la procédure et pendant le calcul de la régression, on voit apparaître sur l'écran l'écart-type des résidus (par modèle, par demi-heure et pour les deux

prévisions). Ces écarts ne doivent pas dépasser le volume du trafic (par demi-heure) du point considéré.

A la fin de la procédure, vérifier si certains coefficients ne sont pas trop forts (supérieurs à 1,5) ou trop faibles (inférieurs à - 1 dans le cas des modèles sans contrainte).

NOTA : Cette étape peut n'être effectuée que par sondage (ce n'est pas le cas de l'étape B6).

C3-2 Fichiers à préparer par l'utilisateur

1) Fichier "débits" du point de prévision
   Fichier (s) "débits amont(s)"
   Ces fichiers ont un format commun qui est décrit ci-dessous.

Format:

1) Colonnes 1- année
2) Colonnes 3-4 mois
3) Colonnes 5-6 jour
4) Colonne 7 type de jour (L=lundi, M=Mardi,W=Mercredi, J=jeudi.....)
5) Colonnes 9-10 jour
6) Colonnes 12-13 mois
7) Colonnes 15-16 année
8) Colonne 19 type de jour (0=Dimanche,l=Lundi, 2=Mardi,.....)
9) Colonnes 25-29 total débit (journalier)

10) Colonnes 30-34 débit premier demi-heure (première heure)

11) Colonnes 35-39 débit deuxième demi-heure (deuxième heure) ainsi de suite débit 48ème demi-heure (24ème heure)

2) Trois fichiers d'informations :

POINT1.INF

ligne 1 : nom du fichier "débits" du point de prévision

ligne 2 : unité du temps pour les débits (48 ou 24)

ligne 3 : nombre d'amonts

POINT2.INF

ligne 1 : nom du point

ligne 2 : sens (Y pour Paris-Province et W pour l'autre sens)

AMONT.INF

ligne 1 : nom du fichier du premier amont

ligne 2 : nom du fichier du second amont etc... (voir la troisième ligne du fichier **POINT1.INF**).

Remarque : Ces noms seront repris dans le fichier "SAISIE" utilisé par ATHENA partie prévision.

3) Fichier "SEUILS" (pour la procédure "ANNEUF")

1ère zone : nombre de zéros au-delà duquel on propose l'élimination d'une ligne.

2ème zone : rapport (en pourcentage) de deux débits consécutifs (dans les deux sens du temps) au-delà duquel on propose l'élimination d'une ligne, mais seulement si les débits considérés sont supérieurs à un certain seuil indiqué dans la troisième zone.

3ème zone : débit en dessous duquel le test sur le rapport de deux débits successifs n'est pas

appliqué.

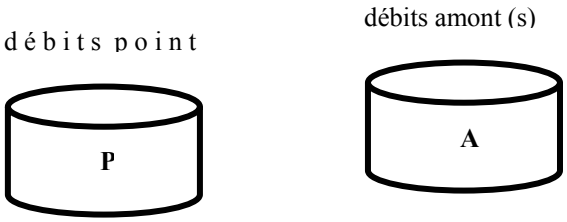
Fichiers de sortie (partie calibrage)

- 24 fichiers de coefficients
- Fichier des résultats des 24 classifications "**SORTIE.OUT**"

Attention tous les fichiers intermédiaires sont supprimés à la fin de la procédure "**CALIB90**".

PREPARATION DES DONNEES

Données de départs :



(un fichier par amont)

0- Sauvegarder les originaux de ces fichiers dans un autre répertoire “TEMPON”

Attention ces fichiers doivent, correspondre à la même période, avoir le même nombre de lignes et être triés de la même façon.

1- Appliquer « selzer » sur le fichier « P »

2- Eliminer dans le fichier "zero.lst" les lignes contenant des "-1" et supprimer les lignes correspondantes dans le fichier "A"

3- Appliquer « selrap » sur le fichier « zero.lst »

4- Consulter le fichier « rapport.lst » pour corriger le fichier « zero.lst »
Si on élimine des lignes dans ce fichier il faut faire de même dans le fichier “A’

5- Trier le fichier « zero.lst » selon mois, année et jour

6- Exécuter « ajout1 » et indiquer: pour le fichier de sortie le fichier « P » et pour le fichier d’entrée le fichier « zero.lst ».

PREPARATION DES DONNEES
suite

8- Appliquer « selzer » sur le fichier « A » (éventuellement modifié)

9- Appliquer « selrap » sur le fichier « zero.lst »

10- Consulter le fichier « rapport.lst » pour corriger le fichier « zero.lst »

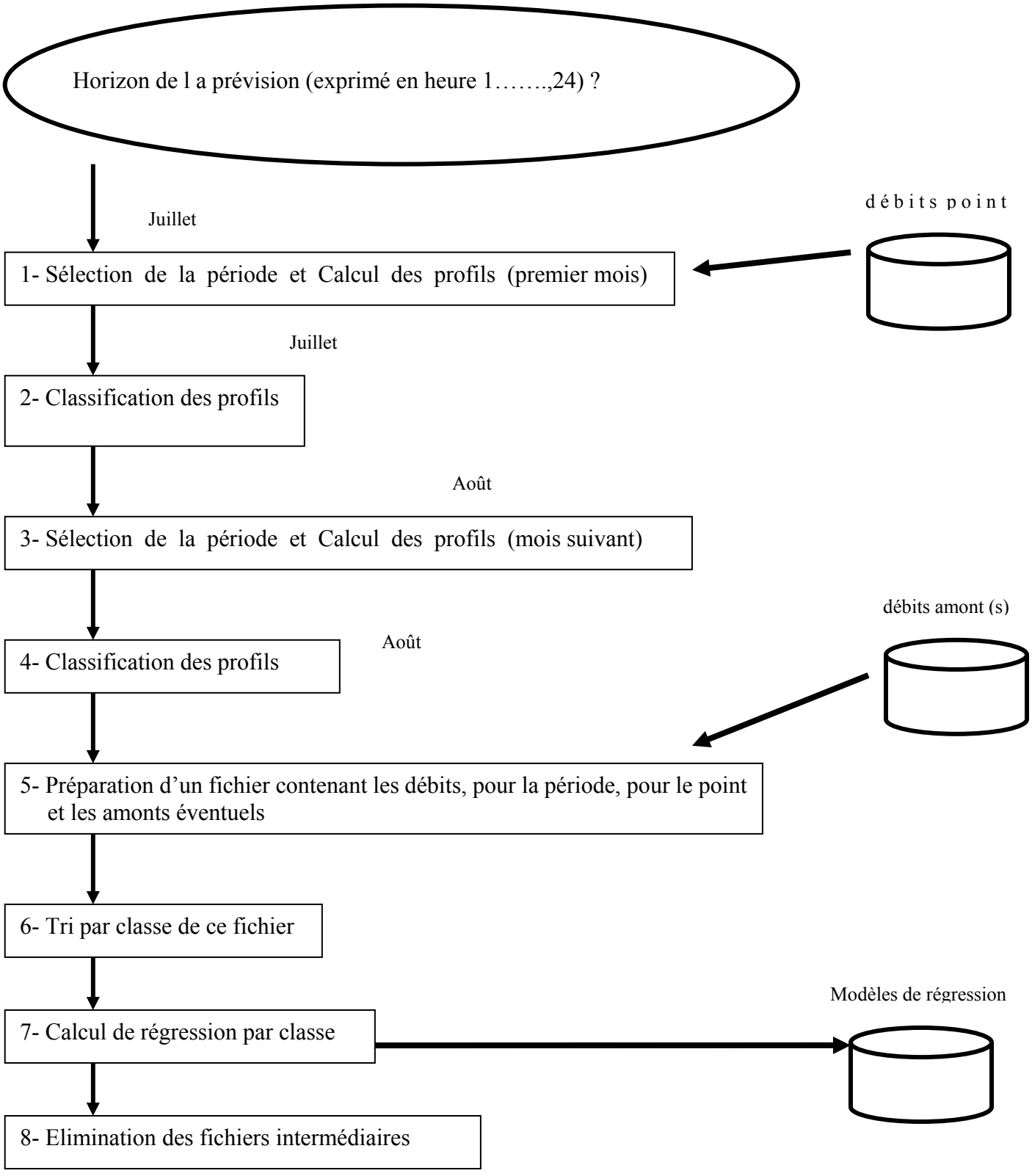
11- Exécuter "ajout1" et indiquer: pour le fichier de sortie le fichier “A” et pour le fichier d’entrée le fichier « zero.lst »

12- Trier le fichier « zero.lst » selon mois, année et jour

13- Vérifier que les fichiers "P" et "A" ont le même nombre de lignes. Ces fichiers doivent être comparables à leurs originaux du répertoire "TEMPON"

En cas d’erreur il faut reprendre les fichiers du répertoire “TEMPON” et retourner à l’étape 1

CALIB 90


Horizon de l a prévision (exprimé en heure 1……..,24) ?
Juillet
1- Sélection de la période et Calcul des profils (premier mois)
débits point
Juillet
2- Classification des profils
Août
3- Sélection de la période et Calcul des profils (mois suivant)
débits amont (s)
Août
4- Classification des profils
5- Préparation d'un fichier contenant les débits, pour la période, pour le point et les amonts éventuels
6- Tri par classe de ce fichier
Modèles de régression
7- Calcul de régression par classe
8- Elimination des fichiers intermédiaires